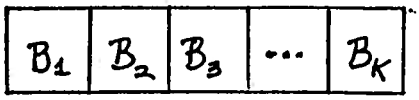
Simple Probability Facts



partition (strata)

$A = AB_1 \cup AB_2 \cup AB_3 \cup \cdots \cup AB_K$

$P(A) = P(AB_1) + P(AB_2) + \cdots + P(AB_k)$

law of total probability.

Defn 1. Conditional prob of A given B $\equiv P(A|B) \equiv \frac{P(AB)}{P(B)}$ (if $P(B) > 0$).

Defn 2. A is independent of B if $P(A|B) = P(A)$.

Thm 1. $P(AB) = P(A|B)\,P(B)$ always.

Thm 2. $P(AB) = P(A)\,P(B)$ if independent.

Why does this definition make sense?

Consider the equally likely case.

$P(A|B) = \frac{N_{AB}/N_S}{N_B/N_S} = \frac{N_{AB}}{N_B}$

Example 1.

$P(\text{1st two cards are diamonds}) = P(D_1 D_2) =$

$= \frac{N_{D_1D_2}}{N_S} = \frac{13 \cdot 12}{52 \cdot 51}$ ← hard

Note: $= P(D_1)P(D_2|D_1) = \frac{13}{52} \cdot \frac{12}{51}$ ← easy (Revisualization)

Fundamental Fact. Conditional probability allows us to revisualize the problem (to make it simpler).

(Rethink the example.) or $= \frac{13}{52} \cdot \frac{12}{51}$

Example 2.

without replacement:

$P(RRWW) = \frac{6}{15}\frac{5}{14}\frac{9}{13}\frac{8}{12}$

$P(RWRW) = \frac{6}{15}\frac{9}{14}\frac{5}{13}\frac{8}{12}$

with replacement:

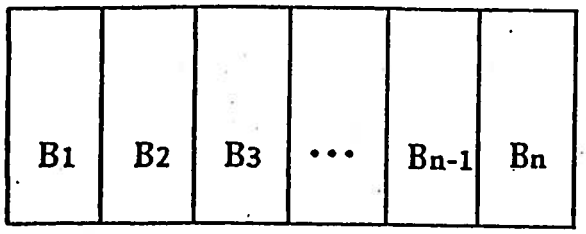
$P(RRWW) = \frac{6}{15}\frac{6}{15}\frac{9}{15}\frac{9}{15}$

$P(RWRW) = \frac{6}{15}\frac{9}{15}\frac{6}{15}\frac{9}{15}$

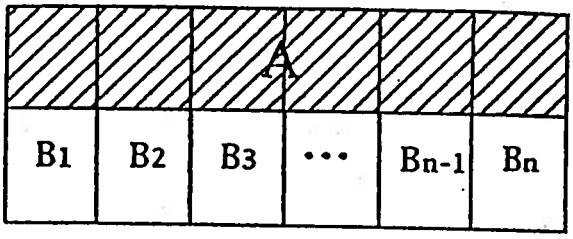
6 R
9 W

$N = 15$
$p = 6/15$

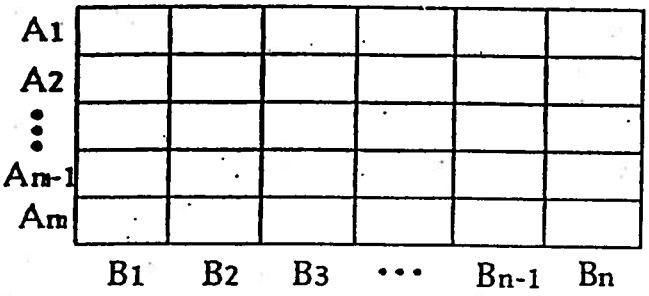
partition
strata



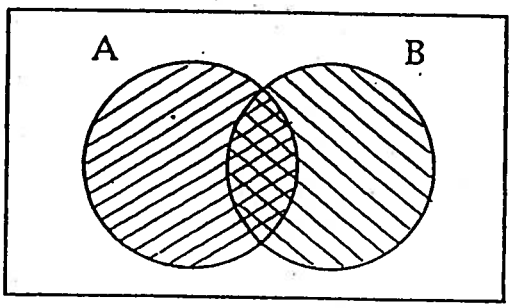
(a)



(b)



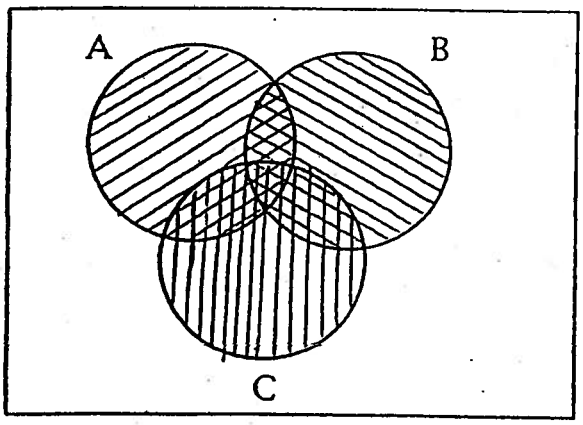
(c)



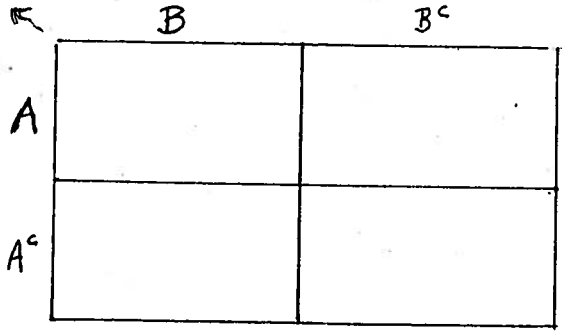
(d)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

or



(e)



(f)

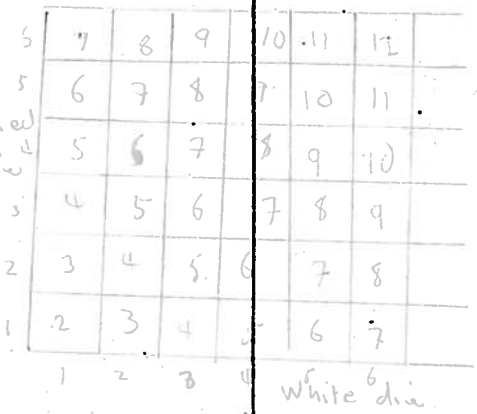
The Game of Craps (Hoyle)

Step 1: Roll two dice. The total T is the "point".
If $T=7$ or $T=11$, then you Win.
If $T=2$ or $T=12$, then you Lose.

Step 2: If $T=$(anything else), then continue rolling until either $T=7$ (then you lose) or $T=$"point" (then you win).

Let W denote the event that you Win.
Let $B_i \equiv [T=i]$ for $2 \le i \le 12$.

B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}	B_{12}



Then

$$P(W) = \sum_{i=2}^{12} P(W \cap B_i) = \sum_{i=2}^{12} P(W|B_i)P(B_i)$$

$$\begin{aligned}
&= P(W\cap B_2) && = P(W|B_2)P(B_2) && = 0 \quad (1/36) && = .00000\\
&+P(W\cap B_3) && +P(W|B_3)P(B_3) && +(2/8)\ (2/36) && +.013889\\
&+P(W\cap B_4) && +P(W|B_4)P(B_4) && +(3/9)\ (3/36) && +.027778\\
&+P(W\cap B_5) && +P(W|B_5)P(B_5) && +(4/10)\ (4/36) && +.044444\\
&+P(W\cap B_6) && +P(W|B_6)P(B_6) && +(5/11)\ (5/36) && +.063131\\
&+P(W\cap B_7) && +P(W|B_7)P(B_7) && +1\quad (6/36) && +.166667\\
&+P(W\cap B_8) && +P(W|B_8)P(B_8) && +(5/11)\ (5/36) && +.063131\\
&+P(W\cap B_9) && +P(W|B_9)P(B_9) && +(4/10)\ (4/36) && +.044444\\
&+P(W\cap B_{10}) && +P(W|B_{10})P(B_{10}) && +(3/9)\ (3/36) && +.027778\\
&+P(W\cap B_{11}) && +P(W|B_{11})P(B_{11}) && +1\quad (2/36) && +.055556\\
&+P(W\cap B_{12}) && +P(W|B_{12})P(B_{12}) && +0\quad (1/36) && +.000000
\end{aligned}$$

$$= .50681$$
$$= .507$$

In Las Vegas ---> .49292

Example 3. $X_1 \equiv$ (# on 1st roll)
$X_2 \equiv$ (# on 2nd roll)
$T \equiv X_1 + X_2 =$ (total # on two dice)

$$P(T \le 3 \mid T \ne 7) = \frac{3}{36-6} = \frac{3}{30} = \frac{1}{10}$$

Purpose: To see various pictures of conditional probability. (four)

Thus our definition makes visual sense, equally likely or not.

Example 4.
Pick urn B_1, B_2, B_3 at random.
Then pick ball at random from that urn.

B_1	B_2	B_3
3R 2W	7R 3W	4R 1W

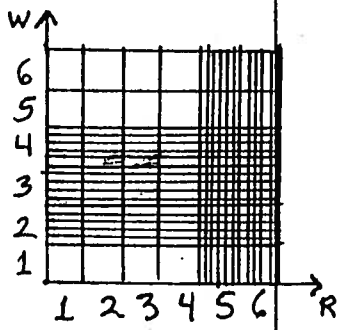
$A \equiv$ [the ball picked is Red].

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_3)P(B_3) = \frac{1}{3}\left(\frac{3}{5} + \frac{7}{10} + \frac{4}{5}\right) = \frac{26}{30}$$

$$P(B_1|A) = \frac{P(AB_1)}{P(A)} = \frac{\frac{1}{3}\frac{3}{5}}{\text{as above}} = \frac{\frac{6}{30}}{\frac{26}{30}} = \frac{6}{26}$$

Bayes Rule. $P(B_i|A) =$ $=$
(do it backwards)

Example 5. $X_1 \equiv$ (# on red die)
$X_2 \equiv$ (# on white die)



$$P(X_1 \ge 5 \text{ and } 2 \le X_2 \le 4) =$$

$$P(X_1 \ge 5)\,P(2 \le X_2 \le 4) =$$

That is: $P(\quad) = P(\quad)P(\quad)$
is how independence typically looks

(Do not use a "mutually exclusive picture.")

EQUIVALENT EXPERIMENTS: Coin flip; Red vs Black card; Even vs Odd dice.

Combanitorics or How to Count

mn Rule:

If m ways to do step I
and n ways to do step II,
then mn ways for two parts.

Ex: Sample 3 times **with replacement:**

ways $= 10^3$

Sample 3 times **without replacement:**

ways $= 10 \cdot 9 \cdot 8$

Values 1 to 10 on ten balls

ways to permute 7 of them =

$n!$ ways to permute n distinct objects. $n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$

$P_k^n \equiv n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ ways to permute k of them.

$C_k^n \equiv \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ subsets of size k.

Pf: $P_k^n = C_k^n \cdot k!$

Ex: (# batting orders of 9 players) $= 9!$

(# batting orders of 26 on roster) $= P_9^{26}$ $= 1.3383 \times 10^{12}$

(# poker hands (of 5 cards)) $= C_5^{52}$ $= 2{,}598{,}960$

(# bridge hands (of 13 cards)) $= C_{13}^{52}$ $= 635{,}013{,}559{,}601$

(# Win, Place, Show with 12 horses) $= 12 \cdot 11 \cdot 10$

(# License plates 3 #'s, 3 letters) $= \binom{6}{3} 26^3 10^3$

If all must be distinct $= \binom{6}{3} P_3^{26} P_3^{10}$

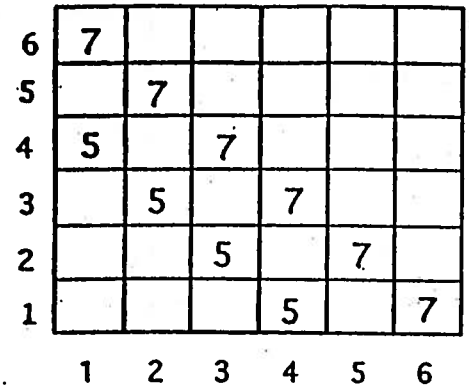
(# permutations of YARD) $= 4!$

(# permutations of MISSISSIPPI) $= \frac{11!}{4!\,4!\,2!}$

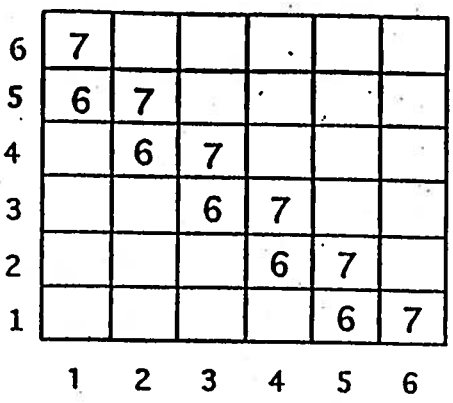
(# permutations of 3★, 3●, 3□, 3△) $= 12!$ if serial numbers

(# ways with like shapes adjacent) $= 4!(3!)^4$

$C_3^5 = \frac{5\cdot 4}{3\cdot 2}$ $C_2^5 = \frac{5\cdot 2}{2\cdot 1}$ $C_1^5 = \frac{5}{1}$ $C_0^5 = 1$ Thus $0! \equiv 1$.



6	7					
5		7				
4	5		7			
3		5		7		
2			5		7	
1				5		7
	1	2	3	4	5	6



6	7					
5	6	7				
4		6	7			
3			6	7		
2				6	7	
1					6	7
	1	2	3	4	5	6

CARDS

13 denominations

	Red		Black	
	♡ Hearts	◇ Diamonds	♧ Clubs	♤ Spades
A=1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
J				
Q				
K				

4 suits
2 colors
$13 \times 4 = 52$ cards

Poker: $C_5^{52} = 2,598,960$

Bridge: $C_{13}^{52} = 635,013,559,601$

Poker

Full house
- 3 of one denomination
- 2 of another denomination

Two pair
- 2 of one denomination
- 2 of another denomination
- 1 of a third

Bridge

4 pts A
3 pts K
2 pts Q
1 pt J
These are the 'honor cards', with their point values.

Poker

$$P(\text{Full House}) = \frac{C_1^{13} C_3^4 C_1^{12} C_2^4 \text{ or } C_2^{13} C_1^2 C_3^4 C_2^4}{C_5^{52}}$$

$$P(\text{Two Pair}) = \frac{C_2^{13} C_2^4 C_2^4 \times \begin{cases} C_1^{44} \\ \text{or } C_1^{11} C_1^4 \end{cases}}{C_5^{52}}$$

Bridge

$T \equiv$ (# of honor cards in random hand of 13 cards)

$$P(T=k) = \frac{C_k^{16} C_{13-k}^{36}}{C_{13}^{52}}$$

Example (3-between) Three face down cards are dealt from a deck. Find the probability the third card falls strictly between the first two. $\frac{52}{52} \frac{48}{51} \frac{44}{50} \cdot \frac{1}{3}$

Expectation (Discrete)

Suppose the basic X-experiment has probability mass function $p(\cdot)$

Thus $p(x) = P(X = x)$.

$\mu \equiv E(X) \equiv EX = (\text{the } \underline{\text{expectation}} \text{ of } X) = (\text{the } \underline{\text{mean}} \text{ of } X)$

$\equiv \sum_{\text{all } x} x\, p(x) = \sum_{\text{all } x} (\text{possibility} \times \text{probability})$

Example.

(a) Bernoulli(p) trial. $\mu =$

(b) Die $\mu =$

(c) Two dice $\mu =$

$Eg(X) \equiv \sum_{\text{all } x} g(x) p(x)$ with μ a special case which measures 'center'.

$\tau \equiv E|X - \mu|$ measures 'spread'; named <u>mean deviation</u>.

$\sigma \equiv \sqrt{E(X-\mu)^2}$ measures 'spread' too; named <u>standard deviation</u>.

$\sigma^2 \equiv E(X-\mu)^2$ named <u>variance</u>.

Example

(a) Bernoulli(p) trial. $\sigma = \sqrt{pq}$ $\tau =$

(b) Die $\sigma = 1.708$ $\tau =$

(c) Two dice $\sigma = 2.415$ $\tau =$

<u>Fact 1.</u> The mean of a sum is the sum of the means. Always!

<u>Fact 2.</u> The variance of a sum is the sum of the variances, for independent

<u>Fact 3.</u> $\sigma^2 = EX^2 - \mu^2$

Examples

$X \cong$ Bernoulli(p) has mean $= \mu = p$ and variance $= \sigma^2 = pq$.

$T \cong$ Binomial(np) has mean $= n\mu = np$ and variance $= n\sigma^2 = npq$.

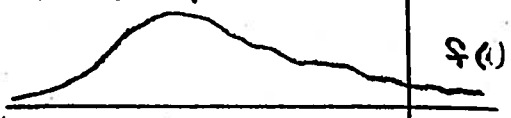
$T \cong$ Poisson(λ) has mean $= \lambda = \nu t$ and variance $= \lambda = \nu t$

$Y \cong$ GeoT(p) has mean $= \frac{1}{p}$ and variance $= \frac{q}{p^2}$.

$W \cong$ NegBiT(p) has mean $= \frac{r}{p}$ and variance $= \frac{rq}{p^2}$.

Expectation (Continuous)

Suppose the basic X-experiment has <u>density</u> $f(\cdot)$



Thus $P(X \in A) = \int_A f(x)\,dx$

for all events.

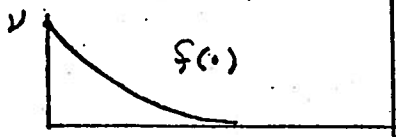
$\mu = EX = (\text{the } \underline{\text{mean}} \text{ of } X) = \int_{\text{all } x} x\, f(x)\,dx$.

$\tau = E|X-\mu| = \int_{\text{all } x} |x-\mu|\, f(x)\,dx$ = <u>mean deviation</u>

$\sigma^2 = E(X-\mu)^2 = \int_{\text{all } x} (x-\mu)^2 f(x)\,dx$ = <u>variance</u> $= EX^2 - \mu^2$

$Eg(X) = \int_{\text{all } x} g(x) f(x)\,dx$.

<u>Facts 1,2,3,</u> Same.



<u>Examples</u>

$Y \cong$ Exponential (rate $= \nu$; or mean $= \theta$)

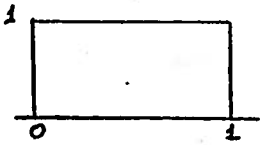
$\mu = \int_0^\infty y\, \nu e^{-\nu y}\,dy = \frac{1}{\nu}\int_0^\infty \nu y\, e^{-\nu y}\,d(\nu y) = \frac{1}{\nu}\int_0^\infty u e^{-u}\,du = \frac{1}{\nu} = \theta$.

$EY^2 = \int_0^\infty y^2 \nu e^{-\nu y}\,dy = \frac{1}{\nu^2}\int_0^\infty (\nu y)^2 e^{-\nu y}\,d(\nu y) = \frac{1}{\nu^2}\int_0^\infty u^2 e^{-u}\,du = \frac{2}{\nu^2} = 2\theta^2$

$\sigma^2 = EY^2 - \mu^2 = 2\theta^2 - (\theta)^2 = \theta^2$.

$\sigma = \theta$.

$X \cong$ Uniform $(0,1)$



$\mu = \frac{1}{2}$ by symmetry

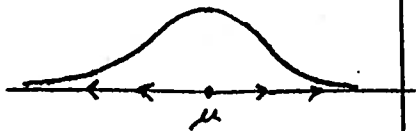
$= \int_0^1 x \cdot 1\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$

$EX^2 = \int_0^1 x^2 \cdot 1\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$.

$\sigma^2 = EX^2 - \mu^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$ $\quad \sigma = \frac{1}{\sqrt{12}}$

$X \cong$ Uniform $(-1,1)$ has $\mu =$ and $\sigma =$

$Z \cong N(0,1)$



$\mu = 0$ by symmetry

$\sigma^2 = EZ^2 - \mu^2 = EZ^2 = \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$

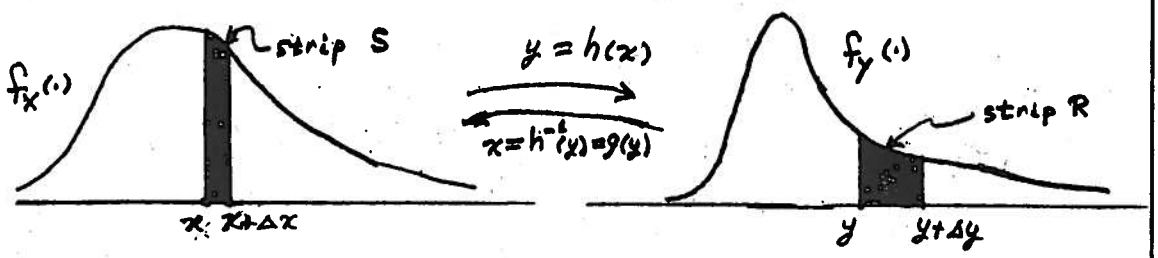
$= \frac{2}{\sqrt{2\pi}}\int_0^\infty z^2 e^{-z^2/2}\,dz =$

$= 1$

CHANGE OF VARIABLE FORMULAS

Transformations (The fundamental idea)

Let $Y = h(X)$ where $h(\cdot)$ is strictly positive and continuous.



The small change from x to $x+\Delta x$ produces the change $\Delta y = h(x+\Delta x) - h(x)$.

Consulting the figure above we see that

$$f_X(x)\,\Delta x \doteq (\text{the area of strip S})$$
$$= P(x \le X \le x+\Delta x) = P(y \le Y \le y+\Delta y)$$
$$= (\text{the area of strip R})$$
$$\doteq f_Y(y)\,\Delta y .$$

That is, $f_Y(y) \doteq f_X(x)\,\frac{\Delta x}{\Delta y}$. Now $\frac{\Delta x}{\Delta y} \to \frac{dx}{dy}\left(= g'(y) = \frac{d}{dy}h^{-1}(y) = \frac{1}{h'(x)} = \frac{1}{h'(h^{-1}(y))}\right)$.

We have thus shown that

$$f_Y(y) = f_X(x)\,\frac{dx}{dy} \quad \text{or} \quad f_{New}(y) = f_{old}(x)\,\frac{d\,old}{d\,new},$$

and this is also true in n dimensions. (That is why we gave this proof.)

Via the DF Let $Y = h(X)$, with $h(\cdot)$ "nice." Let $g \equiv h^{-1}$. Note that

$$F_Y(y) = P(Y \le y) = P(h(x) \le y) = P(X \le h^{-1}(y)) = P(X \le g(y)) = F_X(g(y))$$

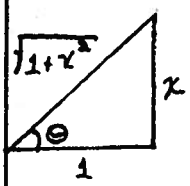
gives

$$f_Y(y) = f_X(g(y))\,g'(y) = f_X(x)\,\frac{dx}{dy} .$$

Math Fact Let $\theta = \tan^{-1}x$, or $x = \tan\theta = \frac{\sin\theta}{\cos\theta}$. Then

$$\frac{d\theta}{dx} = \frac{(\cos\theta)(\cos\theta) - (\sin\theta)(-\sin\theta)}{\cos^2\theta} = \frac{1}{\cos^2\theta} = \frac{1}{(1/\sqrt{1+x^2})^2}$$

$$= 1+x^2,$$



so that

$$\frac{dx}{d\theta} = \frac{1}{1+x^2} .$$

Example (Linear transformations) Let $Y = aX+b$, with $a > 0$.

Then $F_Y(y) = P(Y \le y) = P(aX+b \le y) = P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$

So, $f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) = \frac{1}{a}f\left(\frac{y-b}{a}\right)$.

EXAMPLE 2. (Computer generation of normal rv's) Let Z_1 and Z_2 be independent $N(0,1)$ rv's. Define

(a) $$V = R^2 = Z_1^2 + Z_2^2 \quad \text{and} \quad \Theta \equiv \tan^{-1}(Z_2/Z_1),$$

so that (R,θ) is the polar coordinate representation of (Z_1,Z_2). The Jacobian of the transformation $V = Z_1^2+Z_2^2$ and $\Theta = \tan^{-1}(Z_2/Z_1)$ is

$$\frac{d\text{new}}{d\text{old}} = \left| \begin{array}{cc} 2z_1 & 2z_2 \\ \frac{-z_2}{z_1^2+z_2^2} & \frac{z_1}{z_1^2+z_2^2} \end{array} \right|^{+} = 2. \qquad \text{Thus } \frac{d\text{old}}{d\text{new}} = \frac{1}{2}.$$

Thus

$$f_{V,T}(v,t) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}(z_1^2+z_2^2)\right)/2$$

(b) $$= \frac{1}{2}e^{-v/2}\cdot\frac{1}{2\pi} \quad \text{for } r>0 \text{ and } 0<t<2\pi.$$

Since this factors, that V and Θ are independent rv's. We recognize $\frac{1}{2}e^{-v/2}$ as the Exponential(2) density, so that $V \equiv$ Exponential(2). We also recognize $\frac{1}{2\pi}$ as the Uniform $(0,2\pi)$ density, so that $\Theta \equiv$ Uniform $(0,2\pi)$. That is,

(c) $V = R^2 \equiv$ Exponential(2) and $\Theta \equiv$ Uniform $(0,2\pi)$ are independent.

We now suppose that

(1) U_1 and U_2 denote independent Uniform $(0,1)$ rv's;

in particular, we suppose they are generated by our computer. If $V \equiv -2\log(1-U_1)$, then $P(V \le v) = P(-2\log(1-U_1) \le v) = P(1-U_1 \ge e^{-v/2}) = P(U_1 \le 1-e^{-v/2}) = 1-e^{-v/2}$; so $V \cong Exp(2)$. Clearly $\Theta \equiv 2\pi U_2 \cong Uniform(0,2\pi)$. Thus

(2) $$V \equiv -2\log(1-U_1) \quad \text{and} \quad \Theta \equiv 2\pi U_2 \quad \text{satisfy} \quad f_{U,V}(u,v) = \frac{1}{2\pi}\cdot\frac{1}{2}e^{-v/2}.$$
$$\cong \text{Exponential}(2) \qquad \cong \text{Uniform}(0,2\pi).$$

Transforming back from polar coordinates to, we have

(3) $$Z_1 \equiv \sqrt{-2\log(1-U_1)}\cos(2\pi U_2) \quad \text{and} \quad Z_2 \equiv \sqrt{-2\log(1-U_2)}\sin(2\pi U_2)$$

yields independent $N(0,1)$ rv's $\quad f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{2\pi}\cdot\frac{1}{2}e^{-\frac{z_1^2+z_2^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}\cdot\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2}$.

Thus

(4) $$X_1 \equiv \mu+\sigma Z_1 \quad \text{and} \quad X_2 = \mu+\sigma Z_2$$

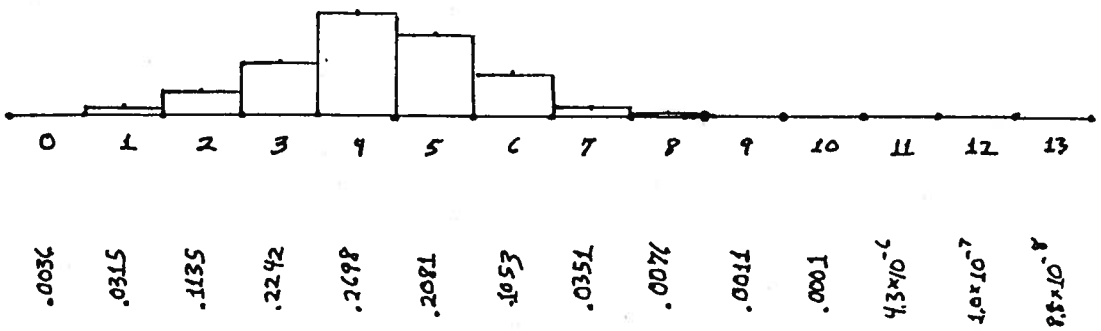
are independent $N(\mu,\sigma^2)$ rv's. By repeating this procedure many times, we can generate as large a sample of $N(\mu,\sigma^2)$ rv's as a simulation study may require. □

GRS | Hypergeometric $(R, W; n)$ distribution

Example ($T \equiv$ # honor cards in bridge hand of 13)

$T = X_1 + X_2 + \cdots + X_{13}$ where each X_i is 1 or 0 according as the i-th card is 'honor' or not.

$$P(T=k) = \frac{C_k^{16} C_{13-k}^{36}}{C_{13}^{52}} \quad \text{for } 0 \le k \le 13$$



These X_i's are dependent rv's. (Even so, we see signs of a CLT.)
(This picuture (of the probability mass function) shows where the probability 'lives'.)

Sampling Without Replacement [Hypergeometric Distribution]

$T = X_1 + \cdots + X_n = (\#\text{ Reds in the sample})$

Since each X_i equals 1 or 0 as i-th draw is R or W.
Dependent repetions of the basic X experiment.

$M = R$	red	success = 1
W	white	failure = 0
N	total	

sample n at random

$p \equiv R/N =$ (fraction of reds)

$$P(T=k) = \frac{C_k^R C_{n-k}^W}{C_n^{R+W}}$$

for each possible k (often $0 \le k \le n$).

$$\text{Mean} = \mu_T = ET = n\,EX = np = n\frac{R}{N}$$

$$\text{Variance} = \sigma_T^2 = \text{Var}[T] = npq\left[1 - \frac{n-1}{N-1}\right]$$

GRS

Binomial(n;p) distribution

Example (Roll a ten sided die 3 times) Outcomes X_1, X_2, X_3

$A = \{3,4,5\}$ will be called success. $p \equiv P(X \text{ is in } A) = \frac{3}{10} = .3$.

$P(X_1 \in A) = \frac{3}{10}$ $\quad P(X_2 \in A) = \frac{3}{10}$

$P(X_1 \in A \text{ and } X_2 \in A) = P(X_1 \in A)P(X_2 \in A) = p \times p = \frac{3}{10} \times \frac{3}{10} = p^2 = \frac{9}{100}$.

OR Let each X_i equal 1 or 0 according as 'success' or 'failure'.

$P(X_1 = 1 \text{ and } X_2 = 1 \text{ and } X_3 = 1) = p^3$

$P(X_1 = 1 \text{ and } X_2 = 1 \text{ and } X_3 = 0) = p^2 q$

$P(X_1 = 1 \text{ and } X_2 = 0 \text{ and } X_3 = 1) = p^2 q$

$T \equiv X_1 + X_2 + X_3 = (\#\text{ of successes in 3 repetitions})$

$P(T = 2) = C_2^3 p^2 q^1 = 3p^2 q$.

$P(T = 3) = C_3^3 p^3 q^0 = p^3$.

Sampling With Replacement [Binomial Distribution]

$T = X_1 + \cdots + X_n = (\#\text{Reds in the sample})$

since each X_i equals 1 or 0 as i-th draw is R or W.
Independent repetitions of the basic X experiment.

R	red = success = 1
W	white = failure = 0
N	total

sample n at random
$p = \frac{R}{N} = (\text{fraction of red})$

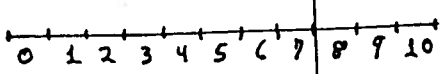
$P(T = k) = C_k^n p^k q^{n-k}$ for $0 \le k \le n$.

as each of the C_k^n possibilities has probability $p^k q^{n-k}$.

Example Basic X-experiment is (headache relief within 20 minutes

$T = X_1 + \cdots + X_{n=10}$

Suppose $T \cong$ Binomial $(n = 10, p = .7)$



if we know $p = .7$.

The rv's $X_1, \ldots, X_n$ are called independent Bernoulli(p) trials.

randomize! Randomize!! RANDOMIZE!!!

Statistics is probability done in reverse.
T is a Binomial (n, p) rv, and we want to estimate p.

Poisson as the Limit of Binomials

GRS

Binomial $(n, p_n) \longrightarrow$ Poisson (λ) when $\lambda_n \equiv n p_n \to \lambda > 0$.

This is especially useful when:
- n is large.
- p_n is small.
- $\lambda_n = n p_n$ is a natural and useful parameter, AND
- n is unobservable.

Examples

Did a Seattle telephone ring up Nordstroms between 9:00 and 10:00 today.
Did a Washington state car have a collision at Broadway and Madison today.
⋮

Let $T_n \cong$ Binomial (n, p_n)

$p_n(k) \equiv P(T_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$ for $0 \le k \le n$.

Let $T \cong$ Poisson (λ)

$p(k) \equiv P(T = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k \ge 0$.

We will show that $p_n(k) \to p(k)$ for each fixed $k \ge 0$, when $\lambda_n \equiv n p_n \to \lambda > 0$.

Well, $p_n(k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \frac{(np_n)^k}{k!} \left(1 - \frac{np_n}{n}\right)^n \frac{1}{\left(1 - \frac{np_n}{n}\right)^k}$

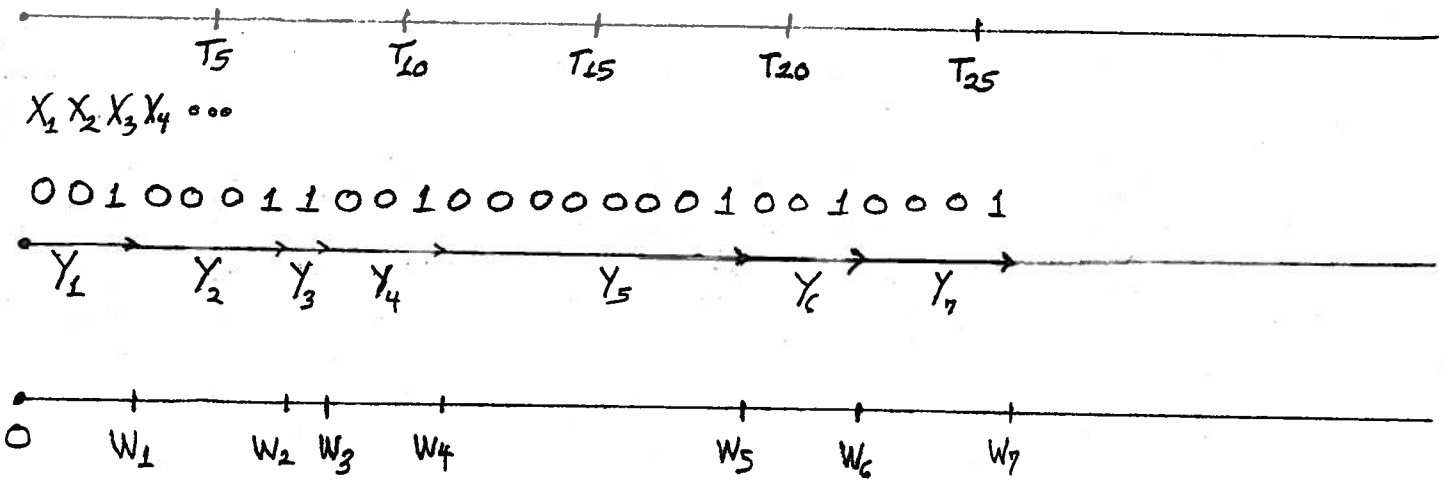
$\to \frac{1 \cdot 1 \cdots 1}{k \text{ times}} \frac{\lambda^k}{k!} e^{-\lambda} \frac{1}{1^k}$

$= \frac{\lambda^k e^{-\lambda}}{k!} = P(\text{Poisson}(\lambda) = k) = P(T = k)$

Comment.

The Poisson (λ) model was born as an approximation to the Binomial. Today it is one of the most important models in its own right. Especially via the Poisson process.

The Bernoulli(p) Process



$T_n \equiv X_1 + \cdots + X_n$

Each $X_i = 1$ or 0

$Y_i \equiv$ (i^{th} interarrival time) until first success

$W_r \equiv$ (r^{th} waiting time) until rth success

Here $X_1, \ldots, X_n, \ldots$ are iid as a Bernoulli(p) rv X; the Bernoulli(p) process.

$T_n \equiv X_1 + \cdots + X_n = $ (# successes in first n turns) $\cong$ Binomial(n,p).

$Y_1, Y_2, \ldots$ are iid as a GeoT(p) rv Y, with $P(Y=k) = q^{k-1}p$ for $k = 1, 2, \ldots$

$W_r \equiv Y_1 + \cdots + Y_r \cong$ NegBiT$(r;p)$ = (# turns till r^{th} success), with $P(W_r = k) = \binom{k-1}{r-1} q^{k-r} p^{r-1} \times p$ for $k \geq r$

Also, $Y_i' \equiv Y_i - 1 \equiv$ (# failures till next success) $\cong$ GeoF(p), with $P(Y=k) = q^k p$ for $k = 0, 1, 2, \ldots$

$W_r' \equiv Y_1' + \cdots + Y_r' \cong$ NegBiF$(r;p)$ = (# failures till r^{th} success), with $P(W_r' = k) = \binom{r+k-1}{r-1} q^k p^{r-1} \times p$ for $k \geq 0$.

Key Fact $[W_r > n] = [T_n < r]$.

Independent Trials

The Basic X-experiment is:

$$X = \begin{cases} 1 & \text{if Success} \\ 0 & \text{if Failure} \end{cases} \quad \text{where} \quad p \equiv P(X=1), \quad q \equiv 1-p = P(X=0).$$

We say $X \cong$ Bernoulli(p). Or, X is a Bernoulli(p) trial.

Examples

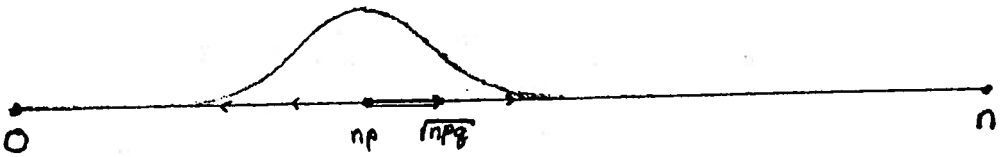
Success/red/1	vs	Failure/white/0	sampling with replacement from an urn.
"6"	vs	"not 6"	when rolling a die.
Headache relief	vs	no relief	twenty minutes after aspirin.
Brand A	vs	Brand B	in a comparison

Let $X_1, X_2, X_3, \ldots$ be a sequence of independent repetitions of the basic expmt.

$$T_n \equiv \sum_{i=1}^{n} X_i = (\text{total number of successes in } n \text{ turns})$$

$$\cong \text{Binomial}(n,p)$$

Thus $P(T_n = k) = \binom{n}{k} p^k q^{n-k}$ for $0 \le k \le n$.



Fact: Sums of independent Binomial rv's are approximately normally distributed, if n is large enough. (They must have the same "p".)

Let $Y_i \equiv$ (# turns after $(i-1)$-st success until i-th success).

We say $Y \cong$ GeoT(p)

So $P(Y=k) = q^{k-1}p$ for $k = 1, 2, \ldots$

Also $1 - F_Y(k) = P(Y>k) = q^k p + q^{k+1}p + \cdots = \frac{q^k p}{1-q} = q^k$ for $k = 1, 2, \ldots$

and $P(Y > i+k \mid Y > i) = q^{k+i}/q^i = q^k = P(Y>k)$

Let $W_r \equiv \sum_{i=1}^{r} Y_i \equiv$ (waiting time until the r-th success)

$\cong$ Negative Binomial$(r, p) \equiv$ NegBiT(r,p).

And $P(W_r = k) = \binom{k-1}{r-1} p^{r-1} q^{k-r} \cdot p$ for $k = r, r+1, \ldots$

Fact: Sum of ind NegBiT rv's again NegBiT (same "p" is required).

Key fact: $[W_r > n] = [T_n < r]$.

Example 2.3.2.(Interviewing job candidates) A personnel manager of a large corporation feels that only 20% of those he interviews would make suitable employees. Let's model his candidate interviews as independent Bernoulli(p) trials X; here X equals 1 or 0 as a candidate is suitable or unsuitable, and $p \equiv P(\text{suitable})$. (Some institutions the manager visits will tend to have a better average candidate than others. Independence would also be violated if good and/or bad candidates tended to be friends, and signed up together to interview. What other potential problems with our model occur to you?) Nevertheless, we consider the following questions in the context of the stated model.

(a). What is the probability the manager's second success occurs in his eighth interview? According to formula, this equals

$$P(W_2 = 8) = \binom{7}{1} p^2 q^6 = 7 \times (.2)^2 \times (.8)^6 = .0734 \doteq .07.$$

(b) He needs to hire 25 candidates. He wonders how many interviews to allow for in his plans. Since he averages 1 out of 5 and needs 25, his first thought is to plan for 125; but then he decides to up this by 10%, and plans for 138. What is the probability he makes 25 offers in 138 interviews? By formula,

$$P(W_{25} \le 138) = P\left(\frac{W_{25} - n\mu}{\sqrt{n}\sigma} \le \frac{138.5 - 25 \times 5}{\sqrt{25}\sqrt{.8/.2^2}}\right)$$

$$= P(Z_n \le 13.5/22.361) = P(Z_n \le .6037) \doteq P(Z \le .6037) = .727.$$

(As a point of comparison, let's approximate this probability another way. Note that the key fact gives

$$P(W_{25} \le 138) = P(T_{138} \ge 25) = P\left(\frac{T_{138} - np}{\sqrt{npq}} \ge \frac{24.5 - 138 \times .2}{\sqrt{138 \times .2 \times .8}}\right)$$

$$= P(Z_n \ge -3.1/4.699) = P(Z_n \ge -.6597) \doteq P(Z \ge -.6597) = .745.$$

Which is closer within the context of the model, .727 or .745? The author would guess that .745 is more accurate, because of the heavily skewed nature of the Geometric(.2) distribution of the interarrival times being summed to give W_{25}. In fact the exact probability can be computed exactly, by formula(s); it is .74131.)

(c) Let's find the smallest n for which n interviews have a 95% chance of finding 25 suitable candidates. Now

$$P(W_{25} \le n) = P(T_n \ge 25)$$

$$= P\left(\frac{T_n - np}{\sqrt{npq}} \ge \frac{24.5 - n \times .2}{\sqrt{n \times .2 \times .8}}\right) = P\left(Z_n \ge \frac{24.5 - .2n}{.4\sqrt{n}}\right)$$

$$\doteq P\left(Z \ge \frac{24.5 - .2n}{.4\sqrt{n}}\right)$$

$$\doteq .95 \quad \text{provided } (24.5 - .2n)/(.4\sqrt{n}) \doteq -1.645.$$

The second equation in the previous line is a quadratic in $\sqrt{n}$; and upon solving it for $\sqrt{n}$ we obtain $\sqrt{n} = 12.835$. Thus $n = 165$ is the desired value (evaluated subject to the normal approximation). □

#1

5.27. A new surgical procedure is said to be successful 80 percent of the time. Suppose that the operation is performed five times and the results are assumed to be independent of one another. What is the probability that

a. all five operations are successful?
b. exactly four are successful?
c. less than two are successful?

#2

5.28. Refer to Exercise 5.27. If less than two operations were successful, how would you feel about the performance of the surgical team? (This type of reasoning will be discussed in Section 5.6.)

#3

5.29. Many utility companies have begun to promote energy conservation by offering discounted rates to consumers who keep their energy usage below certain established subsidy standards. A recent EPA report notes that 70 percent of the island residents of Puerto Rico have reduced their electricity usage sufficiently to qualify for discounted rates. If five residential subscribers are randomly selected from San Juan, Puerto Rico, find the probability that

a. all five qualify for the favorable rates.
b. at least four qualify for the favorable rates.

#4

5.30. Records show that 30 percent of all patients admitted to a medical clinic fail to pay their bills and eventually the bills are forgiven. Suppose that $n = 4$ new patients represent a random selection from the large set of prospective patients served by the clinic. Find the probability that

a. all the patients' bills will eventually have to be forgiven.
b. one will have to be forgiven.
c. none will have to be forgiven.

#5

5.23. Many employers are finding that some of the people they hire are not who and what they claim to be. Detecting falsification of job application information has spawned some new businesses, credential checking services. *U.S. News and World Report* (July 13, 1981), reporting on this problem, notes that, in a two-month period, one service found that 35 percent of all credentials examined were falsified. Suppose that you hired five new employees last week and that the probability that a single employee would falsify information on his or her application form is .35. What is the probability that at least one of the five application forms has been falsified? That two or more forms have been falsified?

#6

5.24. Approximately 10 percent of all thunderstorms tracked over the past five years by the National Oceanic and Atmospheric Administrative Severe Storms Laboratory were "mesocyclones," storms that breed or turn into tornadoes (*Washington Post*, December 7, 1976). Suppose a particular community was subjected to seven thunderstorms and two are judged to be mesocyclonic. What is the probability of observing two mesocyclones out of a total of seven thunderstorms? Of observing two or more? Would you regard two or more as a "rare event," an event that would occur with small probability if in fact the probability of observing a mesocyclonic thunderstorm is only 10 percent?

#7

5.21. The free throw shooting percentages for four NBA basketball players, Alvan Adams, Brad Davis, Kelly Tripucka, and Larry Bird, were given in Exercise 3.39. All scored slightly in excess of 90 percent in the 1984–85 season up to January 7, 1985. Suppose that one of these players received a total of 20 free throws in a single game. For the sake of simplicity, assume that the probability of success on a single free throw is .9 and that the outcomes of the throws represent independent events. What is the probability that the player will make

at least 19 of the 20 free throws?
less than 15 of the 20 free throws?
exactly 90 percent of the 20 free throws?

1 Math Facts

Review of Basic Calculus

Geometric Series The terms of the sequence $a, ar, ar^2, ar^3, \ldots$ are said to *vary geometrically*; here, "a" is called the *first term* and "r" is called the *common ratio*. A specific numerical example with $a = 3$ and $r = \frac{1}{2}$ is given by $3, \frac{3}{2}, \frac{3}{2^2}, \frac{3}{2^3}, \ldots$. The *geometric series* S_n is defined by

$$(1) \qquad S_n \equiv a + ar + ar^2 + ar^3 + \cdots + ar^n,$$

so that $r\,S_n = ar + ar^2 + ar^3 + \cdots + ar^n + ar^{n+1}$. Subtracting rS_n from S_n, most of the terms cancel and we are left with

$$(1-r)S_n = S_n - rS_n = a - ar^{n+1} = a(1 - r^{n+1}).$$

This gives the formula for the *sum of a finite geometric series*, namely

$$(2) \qquad S_n \equiv \sum_{k=0}^{n} a\,r^k = \frac{a}{1-r}\,(1 - r^{n+1}) \qquad \text{for all } n, \text{ and all } r \neq 1.$$

Letting $n \to \infty$ we have $r^{n+1} \to 0$ provided $|r| < 1$. This gives

$$(3) \qquad S_\infty \equiv \sum_{k=0}^{\infty} a\,r^k = \frac{a}{1-r} \qquad \text{if} \quad |r| < 1,$$

as the *sum of an infinite geometric series*. Thus, for example,

$$3 + \tfrac{3}{5} + \tfrac{3}{25} + \tfrac{3}{125} + \cdots = \tfrac{3}{(1-1/5)} = \tfrac{15}{4} = 3.75.$$

Infinite Series Now consider the more general infinite series $\sum_{k=0}^{\infty} a_k x^k$. We say that this series is *absolutely convergent* if $\sum_{k=0}^{\infty} |a_k|\,|x|^k < \infty$ for all $|x|$ less than some number $R > 0$; the largest such R is called the *radius of convergence*. Suppose

$$(4) \qquad f(x) \equiv \sum_{k=0}^{\infty} a_k x^k \qquad \text{is absolutely convergent for } |x| < R$$

One of the basic results of the calculus is that the derivative $f'(x)$ of $f(x)$ is given via *term by term differentiation* to be

$$(5) \qquad f'(x) = \sum_{k=1}^{\infty} k\,a_k\,x^{k-1} \qquad \text{for all} \quad |x| < R.$$

Likewise, any higher order derivative can be found by doing further term by term differentiation for values $|x| < R$ inside the radius of convergence.

Exponential and Logarithmic Series Particularly important infinite series are

$$(6) \qquad e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad \text{for all } x \quad \text{and} \quad \log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k} \quad \text{for all } |x| < 1.$$

(That is, $R = \infty$ and $R = 1$ for these two series, respectively.)

Fundamental Exponential Limit We also know from our calculus class that

$$(7) \qquad \left(1+\frac{a_n}{n}\right)^n \to e^a \quad \text{as } n\to\infty, \text{ whenever } a_n \to a \text{ as } n\to\infty.$$

Exercise 1.4 asks the reader to verify this fact using the expansion of $\log(1+x)$ given above. This will be one of our most used results from the calculus. .

Taylor Series If the kth derivative $f^{(k)}(a)$ exists, then (by *Taylor's theorem*)

$$(8) \qquad f(x) = f(a) + \frac{(x-a)}{1!}f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots + \frac{(x-a)^k}{k!}f^{(k)}(a)$$

$$+\frac{(x-a)^k}{k!}R_k(x), \qquad \text{where } R_k(x)\to R_k(a)=0 \text{ as } x\to a.$$

(We will only need this result for $k=1$ and $k=2$.)

Fundamental Theorem of Calculus Suppose now that $f\geq 0$ is continuous at all but a finite number of points on $(-\infty,\infty)$. Define

$$F(x) = \int_{-\infty}^{x} f(u)\,du \qquad \text{for } -\infty < x < \infty,$$

so that $F(x)$ represents the area that is below f and to the left of x. Then, apart from the finite number of exceptional points,

$$F'(x) = \lim_{h\to 0}\frac{F(x+h)-F(x)}{h} = \lim_{h\to 0}\frac{1}{h}\int_x^{x+h} f(u)\,du$$

$$= \lim_{h\to 0}\frac{\text{the striped area in Figure 1.1}}{h} = \lim_{h\to 0}\frac{h\cdot f(x)}{h} = f(x).$$

Summary When $f\geq 0$ is continuous at all but a finite number of points on $(-\infty,\infty)$, the function

$$(9) \qquad F(x) \equiv \int_{-\infty}^{x} f(u)\,du \qquad \text{has derivative} \qquad F'(x)=f(x)$$

at all x where f is continuous. This is called the *fundamental theorem of calculus.* (Roughly, any F is equal to the integral of its derivative.) [The fancier version of this result is this: If f is integrable, then $F'(x)=f(x)$ is true for almost every x.]

Changing Order of Integration One can integrate a function $f(\cdot,\cdot)$ over the triangular region with vertices $(0,0)$, $(a,0)$, (a,a) either by integrating first over x and then over y, or vise versa. These two integrals will be equal. Thus

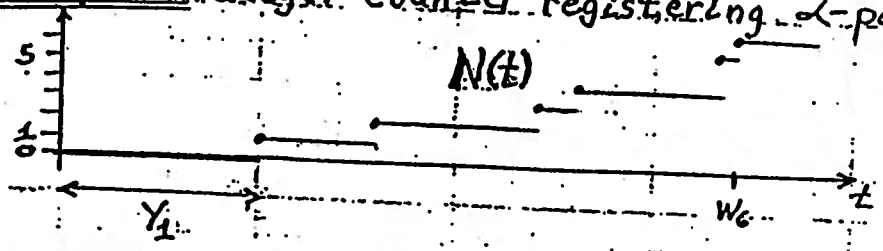
$$(10) \qquad \int_0^a\int_0^x f(x,y)\,dy\,dx = \int_0^a\int_y^a f(x,y)\,dx\,dy.$$

Changing from rectangular to polar coordinates, integration by parts, and general change of variables will all be reviewed in the course of the rest of this section.

Poisson Process $N(t), t \geq 0$.

Looks normal when n large

Example 1: Geiger counter registering α-particles from a radioactive source.



Let $N(t)$ denote the count (the number of successes) at time t.

Let ν denote the intensity, or mean number of counts per unit of time.

Then $\Theta \equiv 1/\nu$ will be seen to be the mean time between counts.

Other examples: Traffic accidents, telephone calls, white corpuscles, defects per foot, ...

Clearly, $N(t)$ has mean νt; and is thus a Poisson(νt) r.v.

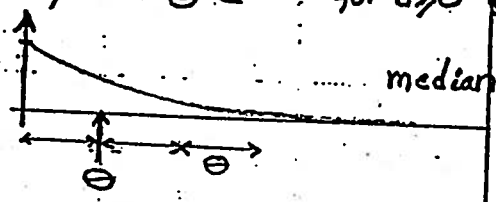
So $P(N(t)=k) = e^{-\nu t}(\nu t)^k/k!$ for $k=0,1,2,\ldots$ That is, $\lambda = \nu t$.

Also $N(t)$ has mean νt, stdev $\sqrt{\nu t}$, and is approximately normal for $\nu t \geq 7$.

Let $Y_1 \equiv$ (the time to the first arrival) $\equiv$ (the first interarrival time).

Then Y_1 has the Exponential(Θ) density $f_Y(t) = \frac{1}{\Theta} e^{-t/\Theta}$ for $t \geq 0$ (or $\nu e^{-\nu t}$).

so Y_1 has mean Θ and stdev Θ.



Fact: $[Y_1 > t] = [N(t) \underset{\text{or } 1}{=} 0]$.

Proof: $1 - F_{Y_1}(t) = P(Y_1 > t) = P(N(t)=0) = e^{-\nu t}$.

Thus $f_{Y_1}(t) = \frac{d}{dt} F_{Y_1}(t) = \frac{d}{dt}[1 - e^{-\nu t}] = \nu e^{-\nu t} = (1/\theta) e^{-t/\theta}$ for $\theta > 0$. □

Now at time Y_1 the process starts over, with Y_2 independent of Y_1.

Thus $Y_1, Y_2, Y_3, \ldots$ are independent Exponential(θ) r.v.'s.

Let $W_k \equiv Y_1 + \cdots + Y_k \equiv$ (the k-th waiting time).

So W_k has mean $k\Theta$ and stdev $\sqrt{k}\Theta$, and is approximately normal for $k \geq 60$.

(This is one of the very worst cases for requiring a large sample size, k.)

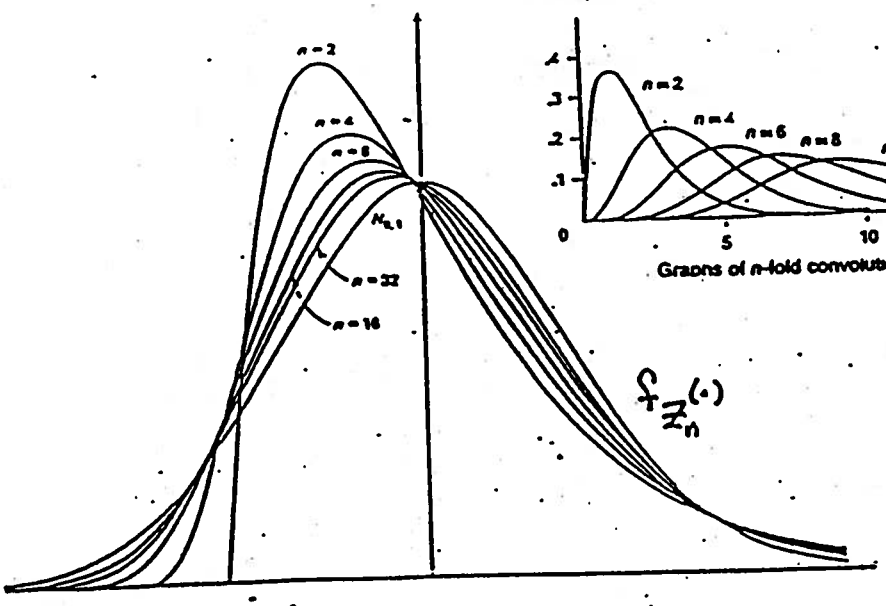
Also W_k has density $f_{W_k}(t) = \frac{t^{k-1}}{\Gamma(k)\Theta^k} e^{-t/\Theta}$ for $t \geq 0$ $\left(\text{or } \frac{\nu^k t^{k-1}}{\Gamma(k)} e^{-\nu t}\right)$,

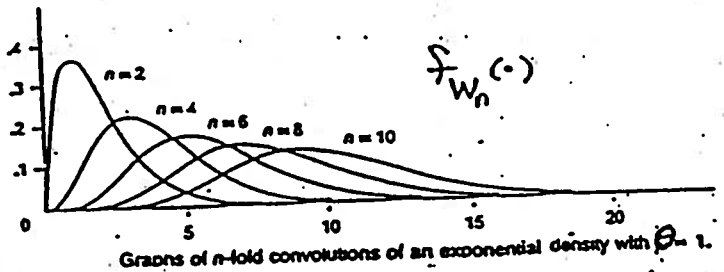
we say that W_k has the Gamma(k, Θ) density.

Fact: $[W_k > t] = [N(t) < k]$.

Proof: $1 - F_{W_k}(t) = P(W_k > t) = P(N(t) < k) = \sum_{i=0}^{k-1} \frac{e^{-\nu t}(\nu t)^i}{i!}$

So $f_{W_k}(t) = \frac{d}{dt} F_{W_k}(t) = \frac{d}{dt}\left[1 - \sum_{i=0}^{k-1} e^{-\nu t}(\nu t)^i/i!\right] = \ldots = \nu^k t^{k-1} e^{-\nu t}/(k-1)!$ □





Graphs of n-fold convolutions of an exponential density with $\theta = 1$.

$W_n \cong \text{Gamma}(n, \theta)$ with $\theta = 1$.
with mean $n\theta$ and stdev $\sqrt{n}\theta$.

$$Z_n \equiv \frac{(W_n - n\theta)}{\sqrt{n}\theta}$$
with mean 0 and stdev 1.

Fact 1: Suppose either: W_{k+1} equals t_0 . or $N(t_0) = k$.
Then there are k arrivals uniformly distributed over $[0, t_0]$. (No proof, but intuitive)
Let $0 < t_1 < t_0$, and let $p \equiv t_1/t_0$. Clearly,
the number in $[0, t_1]$ has a Binomial $(k; p)$ distribution.

Fact 2: If $W_i \cong \text{Gamma}(i, \theta)$ and $W_j \cong \text{Gamma}(j, \theta)$ are independent, then
$W_i + W_j \cong \text{Gamma}(i+j, \theta)$. Proof: $W_i + W_j$ is just like W_{i+j}. □

Fact 3: Now $N(t_1) \cong \text{Poisson}(\nu t_1)$ and $N(t_2) - N(t_1) \cong \text{Poisson}(\nu(t_2 - t_1))$ are independent,
and the sum of these is $N(t_2) \cong \text{Poisson}(\nu t_2)$. Proof: Obvious. □
Generalization: If $X_1 \cong \text{Poisson}(\lambda_1)$ and $X_2 \cong \text{Poisson}(\lambda_2)$ are independent, then $X_1 + X_2 \cong \text{Poisson}(\lambda_1 + \lambda_2)$.

Fact 4: (Lack of memory) $P(Y \geq s+t \mid Y \geq s) = P(Y \geq t)$ for all $s, t > 0$.

• What is probability event happens in t given I count in t_0 one hit in interval

$F(t|\,)$ $P(Y_1 = t \mid N_{t_0} = 1)$

$$\frac{P(Y_1 < t \cap N_{t_0} = 1)}{P(N_{t_0} = 1)} = \frac{P(N_t = 1 \cap N_{t_0} - N_t = 0)}{P(N_{t_0} = 1)} \overset{\text{independence}}{=} \frac{P(N_t = 1)\,P(N_{t_0 - t})}{P(N_{t_0} = 1)}$$

$$= \frac{e^{-\nu t}(\nu t)^1}{1!} \frac{e^{-\nu(t_0 - t)}(t_0 - t)^0}{0!} = \frac{t}{t_0}$$

differentiable to get the cond. density $\frac{1}{t}$

Conditional Randomness of the Waiting Times Suppose we are told that $T_n = m$; that is, that among the first n Bernoulli(p) trials there are exactly m successes. There are C_m^n possible choices as to which m of the n trials resulted in success. We will now show that when $[T_n = m]$ is given, the C_m^n possible choices of the location of the m successes are equally likely. That is, given that m successes have occurred, the location of these success is purely random. Moreover, this is true regardless of the value of the parameter p. This fact will be stated in mathematical notation in (25) below; since all C_m^n possibilities are equally likely, each must have probability $1/C_m^n$.

We will now derive this. Let $x_1, \ldots, x_n$ denote a fixed sequence of 0's and 1's. The desired conditional probability satisfies

$$
\begin{aligned}
&P(X_1 = x_1, \ldots, X_n = x_n \,|\, X_1 + \cdots + X_n = m) \\
&\quad = \begin{cases} 0 & \text{if } x_1 + \cdots + x_n \neq m \\ \frac{P(X_1 = x_1) \cdots P(X_n = x_n)}{P(X_1 + \cdots + X_n = m)} & \text{if } x_1 + \cdots + x_n = m \end{cases} \\
&\quad = \begin{cases} 0 & \\ \frac{p^{x_1} q^{1-x_1} \cdots p^{x_n} q^{1-x_n}}{C_m^n \, p^m q^{n-m}} & \end{cases} \\
&\quad = \begin{cases} 0 & \text{if } x_1 + \cdots + x_n \neq m \\ \frac{1}{C_m^n} & \text{if } x_1 + \cdots + x_n = m \end{cases} \qquad (25) \\
&\quad = \tfrac{1}{C_m^n} \times 1_{\{m\}}(x_1 + \cdots + x_n).
\end{aligned}
$$

(Here, the *indicator function* $1_A(t)$ equals 1 or 0 according as $t \in A$ or as $t \notin A$.)

This result can be used in some interesting ways. The stock market has gradually drifted upward over a period of several years. We might try to model this by supposing that each day there is a constant probability p of the market closing at a higher level than it closed the previous day. That is, whether the daily close is higher or lower than the close of the previous day is modeled as a sequence of independent Bernoulli(p) trials for some unknown $p > \frac{1}{2}$. How can we determine whether or not this model is reasonable? Let $X_1, \ldots, X_n$ denote the outcomes when the basic experiment is repeated on n consecutive business days. Knowing that $X_1 + \cdots + X_n = m$, equation (25) shows that if our model is accurate, then all possible locations of the m successes are equally likely. However, if there was a period of expansion followed by a period of compression (for example), then most of the successes would have come in the first part.

$$P(T_k = i \,|\, T_n = m) = \frac{C_i^m \, C_{k-i}^{n-m}}{C_k^n}$$

m	Red
$n-m$	White
n	Total #
k	Sampled

What is the prob that "i" are Red?

UltraFractal / Fract Int source

n balls are placed into n cells. What is the probability that exactly one cell is empty

- Number of ways to place one ball in n cells

  - n ways to do this for each ball since there are n cells to place each ball.
  - n^n ways to place n balls in n cells

- In order for one cell to remain empty, two balls must be placed in one cell together.

  - 1st ball can go into any cell: n ways
  -

Remaining balls must go into remaining empty cells
Ex: one ball into (n-2) ways
each slot

Gudermannian Function
Hyperbolic trig → Reg Trig

$G(x) \rightarrow$

EX: Set contains n pairs of shoes (2n total shoes) If 2r shoes are chosen at random (with $2r < n$), what is the probability that there will be no matching pairs

- Number of ways to select 2r shoes from 2n shoes: $\binom{2n}{2r}$
- Number of ways to select one shoe from a pair: $\binom{2}{1} = 2$
- Combined number of ways to select 2r shoes from each pair:

$$\binom{2}{1}^{2r} = 2^{2r}$$

- Number of ways to select 2r shoes from n pairs: $\binom{n}{2r}$

(One can only take a single shoe from each pair, thus only n different shoes can be chosen with out selecting any pairs

$$\therefore \quad P = \frac{\binom{n}{2r} \cdot 2^{2r}}{\binom{2n}{2r}}$$

Applications

Example Each August shooting stars (the Perseids) appear at the rate $\nu = .5$ per minute. Let:

$T \equiv N(t) = (\#\text{ shooting stars in } t \text{ minutes}) \cong \text{Poisson}(\lambda = \nu t)$.

$Y \equiv (\text{time till next sighting}) \cong \text{Exponential}(\theta = 1/\nu)$.

Thus $\nu = .5$ and $\theta = 1/\nu = 2$ in this example.

(a) Bart starts watching at midnight. $N(2)$ is the number he sees by 12:02.

$\therefore T \equiv N(2) \cong \text{Poisson}(\lambda = \nu t = .5 \times 2 = 1)$.

Thus the probability Bart sees a shooting star by 12:02 is

$$P(T \geq 1) = 1 - P(T = 0) = 1 - e^{-1} 1^0/0! = 1 - 1/e = .632.$$

(b) The probability Bart sees exactly 3 by 12:04 is (since $T \equiv N(4) \cong \text{Poisson}(\lambda = \nu t = .5 \times 4 = 2)$) given by

$$P(T = 3) = e^{-2} 2^3/3! = .180.$$

(c) The number T Bart sees in an hour is

$$T \equiv N(60) \cong \text{Poisson}(\lambda = \nu t = .5 \times 60 = 30) \doteq N(30, \sqrt{30}^2)$$

Thus the probability Bart sees at least 25 by 1:00 is

$$P(T \geq 25) = P(T \geq 24.5) = P\left(\frac{T-30}{\sqrt{30}} \geq \frac{24.5-30}{\sqrt{30}}\right) \doteq P(Z \geq -1.004) = .842.$$

(d) The time W_{30} it takes Bart to see 30 is

$$W_{30} \equiv Y_1 + \cdots + Y_{30} \cong \text{Gamma}(k = 30, \theta = 2) \cong N(\mu = m\theta = 60, \sigma^2 = m\theta^2 = \sqrt{120}^2)$$

(Look at our graph of f_{W_k} for $k = 32$. This approximation is ok, but not great.)

Let's ask for the probability it takes Bart more than 54 minutes to see 30.

$$P(W_{30} > 54) = P\left(\frac{W_{30} - 60}{\sqrt{120}} > \frac{54-60}{\sqrt{120}}\right) \doteq P(Z > -.5477) = .708.$$

But note that

$$P(W_{30} > 54) = P(N(54) < 30) = P\left(\frac{N(54) - 27}{\sqrt{27}} < \frac{29.5-27}{\sqrt{27}}\right) \doteq P(Z < .481) = .689.$$

(A look at our pictures of Poisson distributions makes us believe this approximation is closer. The true value is .6935...)

(e) Let W_2 denote the time from Bart's first sighting until his third. Then

$W_2 \cong \text{Gamma}(k=2, \theta=2)$ with density $f_{W_2}(t) = \frac{t}{4} e^{-t/2}$ for $t>0$.

Thus

$$P(W_2 > 3) = \int_3^\infty \frac{t}{4} e^{-t/2} dt = -\frac{t}{2} e^{-\frac{t}{2}} - e^{-\frac{t}{2}} \Big|_3^\infty = e^{-\frac{3}{2}}\left(\tfrac{3}{2}+1\right) = .558.$$

same as $P(N_3 < 2) = P(\text{Poisson}(\tfrac{3}{2}) < 2) = e^{-3/2}\left[(\tfrac{3}{2})^0/0! + (\tfrac{3}{2})^1/1!\right]$

(f) Suppose Barts sees 8 in the first 20 minutes. What is the probability that exactly 3 are seen in the first 5 minutes. According to Poisson process Fact 1, this number N satisfies

$M \cong \text{Binomial}(n=8,\ p = \frac{5}{20} = .25)$.

Thus

$$P(M=3) = \binom{8}{3} .25^3 .75^5 = .208.$$

(g) Bart went out at midnight. Given he waited more then 5 minutes for his first sighting, what is the probability he waited more then 8?

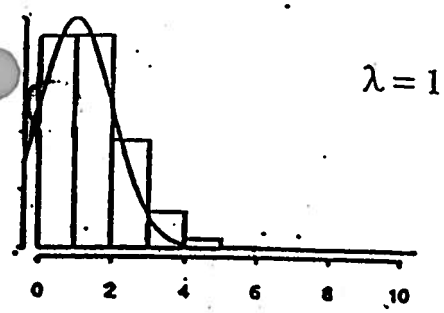
$$P(Y>8 \mid Y>5) = P(Y>3) = e^{-t/\theta}\Big|_{t=3} = e^{-3/2} = .223 \text{ by lack of memory.}$$

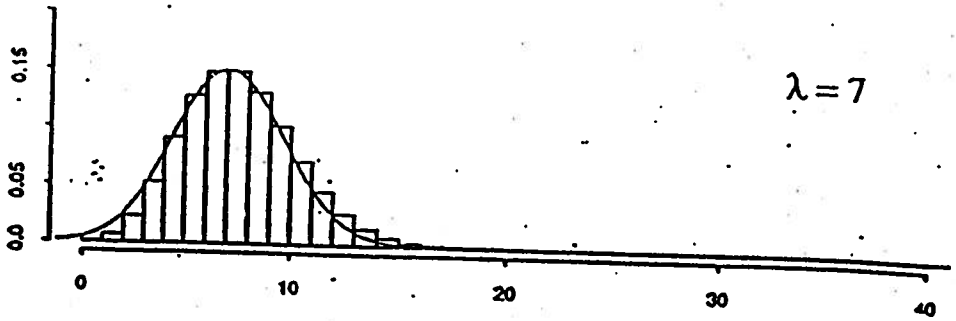
(h) Bart will get bored and quit if he has to wait more then 3 minutes between sightings. What is the probabilty p Bart will see exactly 5 shooting stars and then quit. well

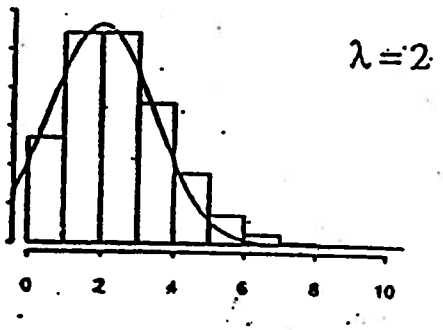
$$p = P(Y_1 < 3)P(Y_2 < 3)P(Y_3 < 3)P(Y_4 < 3)P(Y_5 < 3)P(Y_6 > 3)$$
$$= P(Y<3)^5 P(Y>3) = .777^5 \times .223 = .0632 \quad \text{(see (g))}.$$

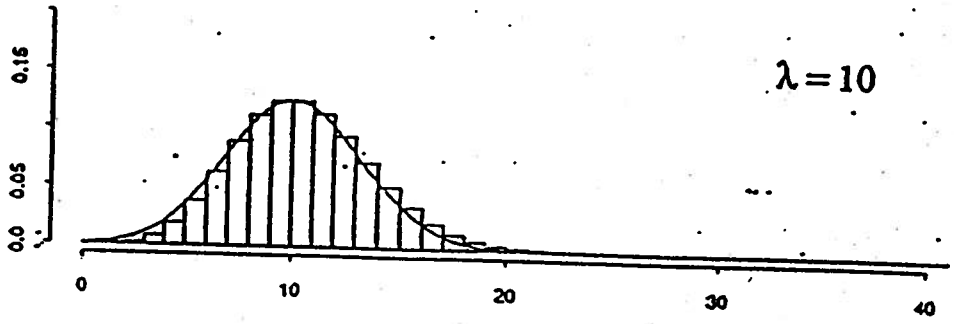
Turns

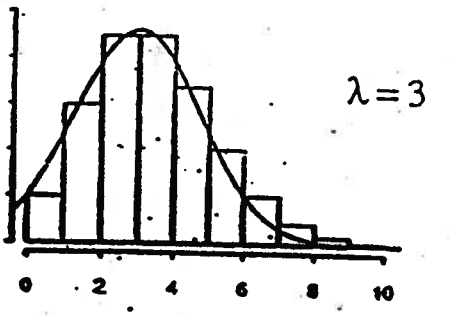
Geometric Distribution

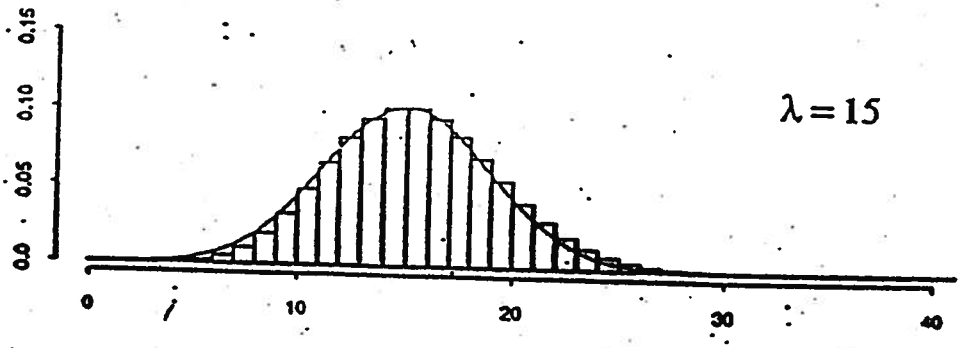


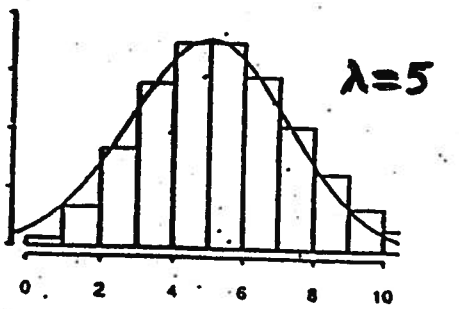


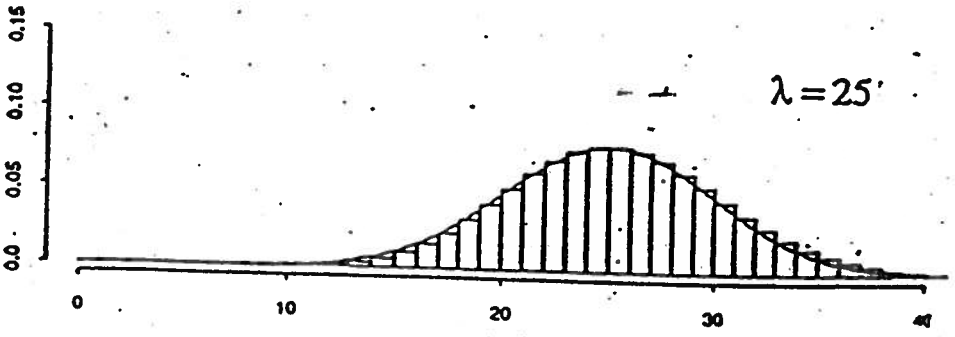












P(Y=k) when Y is Poisson(λ)

QA 188. G65 1996

For Lecture 3

1. A Scrabble set has 100 equal sized chips; some consonant is written on 54 of them, some vowel is written on another 44, and 2 of them are blank. You draw 7 chips at the start of the game. Let T denote the number of consonants in your sample. Evaluate $p(k) \equiv P(T = k)$ for $0 \leq k \leq 7$.
(a) Graph the probability distribution $p(\cdot)$ of T.
(b) Evaluate $\mu_T \equiv ET$ and $\sigma_T^2 \equiv Var[T]$.

Chapter 1 #20 from C and B set intersection / union
Chapter 3 #4,6,12,23,24b from C and B

For Lecture 4

2. Let us regard shucking oysters in search of a natural pearl as independent Bernoulli(p) trials with $p = .04$. A diver must search out 25 pearls per day.
(a) Evaluate $P(T_{590} \geq 25)$ and $P(T_{660} < 25)$, for T_n successes in n attempts.
(b) Let W_{25} denote the number of attempts required to obtain 25 pearls. Sketch the normal approximation to the distribution of W_{25}. Will the normal approximation be reasonably adequate? Evaluate $P(W_{25} \leq 590)$ and $P(W_{25} > 660)$.
(c) What is the probability that none of the eight divers employed by the company will have exceeded 590 attempts to make their quota of 25 pearls?
(d) Some of the oysters are randomly selected and brought to a nearby gift shop. A tourist is willing to pay for eight oysters in hopes of finding at least one pearl. What is the probability this tourist finds none?
(e) Another tourist intends to shuck oysters until she finds a pearl. What is the probability that she will find success within eight tries?

3. (a) On a day when Rodney Rocket is serving well (his first serve is landing in bounds 80% of the time) he decides to put a bit more speed than usual on his second serve (it now lands in only 60% of the time). What is the probability Rocket "double faults" with both serves bad?
(b) On a day when Rocket is serving poorly (his first serve is landing in bounds only 40% of the time) he is a bit more careful than usual on his second serve (it now lands in 95% of the time). What is the probability Rocket double faults?
(c) Let Y denote the number of turns until Rocket's first double fault. Compute μ_Y and σ_Y in both of the models (a) and (b).

4. What is the probability of getting at least 6 honor cards in at least 20 of 100 hands of bridge?

5. Three brothers GR, Bart, and Matthew go to Las Vegas, and each plays a different game. GR pays \$4 each time he rolls a single die, and he is rewarded with a dollar for each spot on the face that turns up. Bart pays \$7 to roll a die until a "one" appears, and he is rewarded with a dollar for each turn until this happens. -(If the first "one" occurs on his 8th turn, then Bart receives \$8.) Matthew pays \$3 to draw a ball from an urn containing three balls numbered 1, 2, and 3, and he is rewarded with a dollar value equal to the serial number he draws.
(a) Let X, Y, and Z denote the winnings (payoff minus cost) of GR, Bart, and Matthew on one turn of their respective games. Compute the mean and standard deviation of X, Y, and Z.
(b) Each brother plays his game ten times. Let T denote the total winnings of the three brothers combined. Compute the mean and standard deviation of T. Evaluate the probability $P(T \geq 0)$ that the brothers do not lose money.
(c) Sketch a graph of the distribution of T, and put labels on the horizontal axis.

For Lecture 5

6. Calls at the 911 number in Our Town occur throughout the "nighttime period" from midnight until 3:00 am at the rate of 12 calls per hour.
(a) Let N denote the number arriving between 1:00 and 1:30. Evaluate $P(N \leq 2)$.
(b) What is the probability that the next call is more than 125 minutes away?
(c) Police call it a "dangerous situation" if the midnight to 3:00 am operators must deal with at least 50 calls. What is the probability (on any specific night) that such a dangerous situation occurs?
(d) Find the probability of no dangerous situations during a seven day week?
(e) Last Tuesday was rather unusual, as only 6 calls were received. Given that this happened, what is the probability they all arrived between midnight and 1:00 am?

7. (Mine explosions in England) In a series of mine explosions in England, the time interval between the first and fiftieth explosions was 7350 days (say).
(a) State what mathematical model you will use to describe this situation. Give the value of your estimate of each parameter of that model. Let p̂robability be used to denote probabilities computed under your model, in which you use the estimated values of the parameters.
(b) What is the m̂ean number of explosions for the extended period beyond this data from day 7350 through day 8200?
(c) Find the p̂robability of two or less explosions from day 7350 through day 8200?
(d) What is the p̂robability of fifteen or more such disasters in the next ten years?
(e) What is the p̂robability that the fifty-first mine explosion occurs after day 7450.
(f) What is the p̂robability that the fifty-third explosion occurs after day 7650?
(g) Find the p̂robability of waiting more than 4500 additional days for the thirtieth aditional mine failure?
(h) Given that the fifty-first explosion occurred after day 7450, find the p̂robability it occurred after day 7550?
(i) What is the p̂robability that the time between each of the next six explosions is never more than one hundred days?
(j) What is the p̂robability that the times between each of the next six explosions exceeds one hundred days exactly twice?
(k) Given that 3 additional mine explosions occur in the next year, what is the p̂robability of two in the next two months?
(l) Comment: Mine explosions typically occur due to the buildup of methane gas. Note question (h) again in this regard, as it highlights a weakness of the Poisson process model for this particular data set. Still, the model should do a reasonable job of describing average behavior over an extended period of time. If the disasters in question in this problem were airplane crashes, the model would would seem to fit even better—particularly in regard to (h).

"PLANNED" Syllabus for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29	L1...Ch1: Probability, conditional probability, independence, and rvs
Mar 31	L2...Sec2.1–2.3$^-$: E(·), mean, variance, and (especially) transformations (no mgf)
Apr 01	Q....HW and some catchup lecturing
	L3...Sampling with and without replacement
.....	
Apr 05	L4...Sec3.1–3.2(Handout): Bernoulli trials; Geometric, NegBinomialT (and Hypergeometric and Discrete Uniform)
	L5...Sec3.3(Handout): Poisson process, Poisson, Exponential, Gamma (and Beta and Uniform)
	Q ....HW, and finish Poisson process
	L6...Sec2.3: Moment generating function (mgf)
.....	
Apr 12	L7...More mgf, and CLT pics
	L8...Ch4: Joint, marginal, and conditional distributions
	Q....Sec4.1: Conditionals and convolution
	L9...Sec4.2: Conditional mean and variance
.....	
Apr 19	L10..Sec4.2 Independence and Expectation
	L11..Other Ch4 material
	Q....MIDTERM I
	L12..Discuss Midtem, and begin anew
.....	
Apr 26	L13..Sec4.5: Covariance
	L14..Sec5.1–5.2: Properties of $\bar{X}$ and S^2
	Q....
Apr 30	L15..Sec4.5: Bivariate Normal Distribution
.....	
May 03	L16..Sec5.3: Derivation of Student-T and Fisher/Snedecor/varianceRatio-F
	L17..Handout: Moments via matrices
	Q....
	L18..Sec7.1–7.2:Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)
.....	
May 10	L19..MLE: Three illustrative MLE examples
	L20..Sec6.1–6.2: Sufficiency
	Q....Work HW on MLE
	L21..Completenes, Exponential families, ancillarity
.....	
May 17	L22..Bayes estimators
	L23..Bayes examples
	Q....MIDTERM II
	L24..Discuss Midterm, and begin anew
.....	
May 24	L25..Cramer–Rao inequality
	L26..Finish Cramer–Rao
	Q....
	L27..Hypothesis tests
.....	
May 31	MEMORIAL DAY HOLIDAY
June02	L28..Likelihood ratio tests
	Q....Work examples of LR-tests
June04	L29..Location and scale models
.....	
June09	FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....	

Planned Homework for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29	L1...Chapter1: #[2,5,11,13,14,25,33,38; do not be formal; should be easy];#17,18,21,23,24
Mar 31	L2...Chapter1: #44,53; Chapter2: #4,8c,17,19,24
Apr 01	Q....
	L3...Chapter2: #20; Chapter3: #4,6,12,23,24b; Handout: #1
.....	
Apr 05	L4...Handout: #2–5
	L5...Handout: #6–7
	Q ...
	L6...
.....	
Apr 12	L7...Chapter2: #9,16, (read 11); Chapter3: #22cd
	L8...Chapter2: #33abc; Chapter3: #19,22ab
	Q....
	L9...Chapter2: # 26ab, 28ab, 30ad; Chapter3: #24c
.....	
Apr 19	L10..
	L11..
	Q....
	L12..
.....	
Apr 26	L13..
	L14..
	Q....
Apr 30	L15..
.....	
May 03	L16..
	L17..
	Q....
	L18..
.....	
May 10	L19
	L20
	Q....
	L21..
.....	
May 17	L22..
	L23..
	Q....
	L24..
.....	
May 24	L25..
	L26..
	Q....
	L27..
.....	
May 31	MEMORIAL DAY HOLIDAY
June02	L28..
	Q
June04	L29..
.....	
June09	FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....	

M 2:30 – 6 W 2:30 → 6
T 3 – 6 Th 11:30 → 4

Moment Generating Function (MGF)

GRS

$= \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$

DEFN 1. $M(t) \equiv M_X(t) \equiv E(e^{tX})$, for real t, is called the mgf of X.

$\phi(t) \equiv \phi_X(t) \equiv E(e^{itX})$, for real t, is called the characteristic function,
$= \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$ (always exists)

FACTS

(1) $M_{aX+b}(t) = E\,e^{t(aX+b)} = e^{tb}\,E e^{(at)X} = e^{tb} M_X(at)$.

$X = aX + b$ ↑ linear transf

(2) $M_{X+Y}(t) = M_X(t) M_Y(t)$ for X and Y independent

Since $M_{X+Y}(t) = E\,e^{t(X+Y)} = E(e^{tX}e^{tY}) \underset{\text{by ind}}{=} E e^{tX} E e^{tY} = M_X(t) M_Y(t)$. (for convenience)

(3) (Uniqueness theorem) $\phi_X(\cdot)$ determines $P_X(\cdot)$, and conversely.
So does $M_X(\cdot)$ (at least, if it's finite in a neighborhood of 0).

(4) (Cramér-Lévy continuity theorem) Let Z_n and Z be rv's.
If $M_{Z_n}(t) \to M_Z(t)$ for all t in a neighborhood of 0, then
$Z_n \to_d Z$ $F_{Z_n}(z) = P(Z_n \le z) \to P(Z \le z) = F_Z(z)$ (provided $F_Z(\cdot)$ is continuous at z).
[$\phi_{Z_n}(t) \to \phi_Z(t)$ for all t iff $Z_n \to_d Z$.]

EXAMPLES

(a) [Bernoulli(p)] $M(t) = e^{t\cdot 1}\cdot p + e^{t\cdot 0}\cdot(1-p) = 1 + p(e^t - 1)$.

(b) [Binomial(n,p)] $M(t) = [1 + p(e^t - 1)]^n$ (it's the sum of n independent Bernoulli)

(c) [Poisson(λ)] $M(t) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$.

(d) [Exponential(rate=λ)] $M(t) = \int_0^{\infty} e^{ty}\lambda e^{-\lambda y}\,dy = \frac{\lambda}{\lambda - t}\int_0^{\infty} (\lambda - t) e^{-(\lambda - t)y}\,dy = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.
$= \frac{1}{1 - \theta t}$

(e) [Gamma(r,λ)] $M(t) = \left(\frac{\lambda}{\lambda - t}\right)^r = \left(\frac{1}{1 - \theta t}\right)^r$ for $t < \lambda$.

(f) [$N(\mu, \sigma^2)$] $M(t) = e^{\mu t + \sigma^2 t^2/2}$

since $M_{N(0,1)}(t) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = e^{t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2}\,dz = e^{t^2/2}\cdot 1 = e^{t^2/2}$.

(g) [GeoF(p)] $M(t) = \sum_{k=0}^{\infty} e^{tk} p q^k = p\sum_{k=0}^{\infty} (qe^t)^k = \frac{p}{1 - qe^t}$ for $t < \log\frac{1}{q}$.

WHY CALLED THE MGF? Now $e^{tX} = 1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$. "Thus"

$M(t) = 1 + tEX + \frac{t^2}{2!}EX^2 + \frac{t^3}{3!}EX^3 + \cdots$. So $\frac{d^k}{dt^k} M(t)\Big|_{t=0} = EX^k$.

Suppose that X is $(0, \sigma^2)$. Then

(a) $M_{X/\sqrt{n}}(t) = M_X\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t}{\sqrt{n}}\cdot 0 + \frac{t^2\sigma^2}{2n} + \frac{t^3 EX^3}{6\,n^{3/2}} + \cdots = \left(1 + \frac{\frac{\sigma^2 t^2}{2} + \frac{t^3 EX^3}{6\sqrt{n}} + \cdots}{n}\right)$

Tends to 1 ... more difficult way to find moments

CLT Let $X_1, \dots, X_n$ be iid $(0, \sigma^2)$. Then $Z_n \equiv \frac{1}{\sqrt{n}} \bar{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \to_d N(0, \sigma^2)$.

Proof. $M_{Z_n}(t) = E e^{t Z_n} = E e^{t \frac{1}{\sqrt{n}} \sum X_i} = E e^{\sum \left(\frac{t}{\sqrt{n}} X_i\right)} = \prod_{1}^{n} E e^{\frac{t}{\sqrt{n}} X_i} \underset{\text{by ind}}{=} \prod_{1}^{n} E e^{\cdot}$

$= \prod_{i=1}^{n} M\left(\frac{t}{\sqrt{n}}\right) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$

$= \left[1 + \frac{\sigma^2 t^2}{2} + \frac{1}{\sqrt{n}} \frac{t^3 E X^3}{6} + \cdots}{n}\right]^n \equiv \left[1 + \frac{a_n}{n}\right]^n$ where $a_n \to a \equiv \frac{\sigma^2 t^2}{2}$

$\to e^a = e^{\frac{1}{2}\sigma^2 t^2} = M_{N(0,\sigma^2)}(t)$.

Thus $Z_n \to_d N(0, \sigma^2)$ by Cramér–Lévy continuity theorem. □

(Are you any smarter? How big does n have to be?

PLT Let $X_{n1}, \dots, X_{nn}$ be independent Bernoulli (p_n), where $np_n \to \lambda$.

Then $T_n \cong \text{Binomial}(n, p_n) \longrightarrow_d \text{Poisson}(\lambda)$.

Proof. $M_{T_n}(t) = [1 + p_n(e^t - 1)]^n = \left[1 + \frac{\overbrace{np_n(e^t-1)}^{a_n}}{n}\right]^n \equiv \left[1 + \frac{a_n}{n}\right]^n$

$\to e^a = e^{\lambda(e^t - 1)}$ since $a_n \equiv np_n(e^t - 1) \to \lambda(e^t - 1)$

$= M_{\text{Poisson}(\lambda)}(t)$. Poisson Limit Theorem

Thus $T_n \to_d \text{Poisson}(\lambda)$ by Cramér-Lévy continuity theorem. □

EXAMPLE (a) The sum of independent normals is normal. That is, independent $X_1 \cong N(\mu_1, \sigma_1^2)$ and $X_2 \cong N(\mu_2, \sigma_2^2)$ gives $X_1 + X_2 \cong N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
(But we knew this anyway, from the CLT. Why? Think probabilistically!)
(b) The sum of independent gammas is gamma. That is, independent $X_1 \cong \text{Gamma}(r_1, \lambda)$ and $X_2 \cong \text{Gamma}(r_2, \lambda)$ gives $X_1 + X_2 \cong \text{Gamma}(r_1 + r_2, \lambda)$.
(c) Same for ...

intuitively add many things → normal add more → more normal

Proof

(Normal) $M_{X_1 + X_2}(t) = M_{X_1}(t) M_{X_2}(t) = e^{\mu_1 t + \sigma_1^2 t^2/2} e^{\mu_2 t + \sigma_2^2 t^2/2} = e^{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2} = M_{N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}$

(Gamma) $M_{X_1 + X_2}(t) = M_{X_1}(t) M_{X_2}(t) = \left(\frac{\lambda}{\lambda - t}\right)^{r_1} \left(\frac{\lambda}{\lambda - t}\right)^{r_2} = \left(\frac{\lambda}{\lambda - t}\right)^{r_1 + r_2} = M_{\text{Gamma}(r_1 + r_2, \lambda)}(t)$. □

EXERCISE Prove (a) Normal and (b) Gamma via the convolution formula. (Multiplying is easy. Convolving is not.)

12.4.2004

GRS

BETA RVs AND ORDER STATISTICS

Let $U_1, \ldots, U_n$ be independent Uniform(0,1) rv's. ordered left to right

Let $0 \equiv U_{n:0} \leq U_{n:1} \leq \cdots \leq U_{n:n} \leq U_{n:n+1} \equiv 1$ be the <u>order statistics</u>

Let $T_n(t) = \sum_{i=1}^{n} 1_{[0,t]}(U_i) = (\#\text{ of } U_i\text{'s that are } \leq t)$ $\sim Bi(n, p=t)$

for fixed t counts num. of observ. in interval 0 to t.

<u>KEY FACT</u> order statistic point counts $[U_{n:i} > t] = [T_n(t) < i]$

i.e. The Binomial and the Beta are closely related.

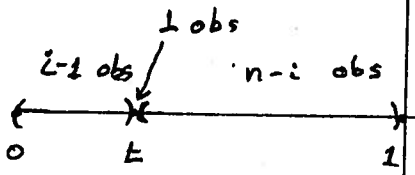
<u>Note</u> $1 - F_{U_{n:i}}(t) = P(Bi(n,t) < i) = \sum_{k=0}^{i-1} \binom{n}{k} t^k (1-t)^{n-k}$.

Differentiate this to find $f_{U_{n:i}}(\cdot)$. □

Sum telescopes

"OR"

$$f_{U_{n:i}}(t) = \binom{n}{i-1, 1, n-i} t^{i-1} (1-t)^{n-i}$$



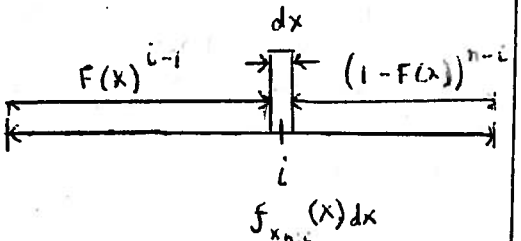
$$= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} t^{i-1} (1-t)^{(n-i+1)-1}$$

We saw last time this rv has: mean $= \frac{i}{n+1}$, var $= \frac{1}{n+2} \frac{i}{n+1} \left(1 - \frac{i}{n+1}\right)$

Meanwhile: $\frac{1}{n}$ Binomial$(n, \frac{i}{n+1})$ has: mean $\frac{i}{n+1}$, var $= \frac{1}{n} \frac{i}{n+1} \left(1 - \frac{i}{n+1}\right)$

Probability of X_i u_i

$$f_{X_{n:i}}(x)\,dx = \binom{n}{i-1}\binom{n-(i-1)}{1}\binom{n-i}{n-i} F(x)^{i-1} f(x)\,dx \cdot (1 - F(x))^{n-i}$$

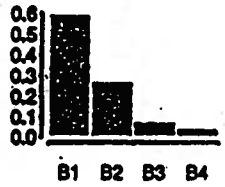


multinomial coefficient

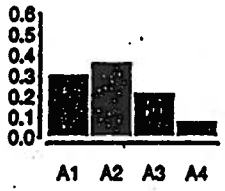
$$= \frac{n!}{(i-1)!\,1!\,(n-i)!} = \binom{n}{i-1, 1, n-i}$$

$$\binom{n}{i-1}\binom{n-(i-1)}{1}$$

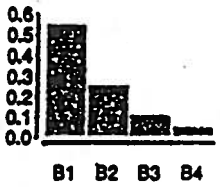
AGE (A1-A4) vs PATRONIZED FAST FOOD CHAIN (B1-B4)



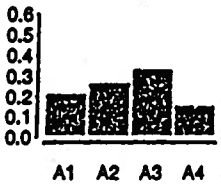
A1: 16-21 [.25]



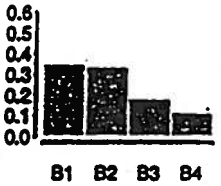
B1: McDonald's [.478]



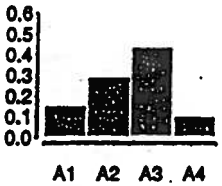
A2: 21-30 [.32]



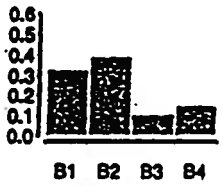
B2: Burger King [.306]



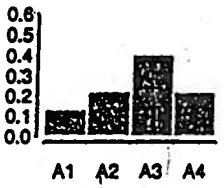
A3: 30-49 [.304]



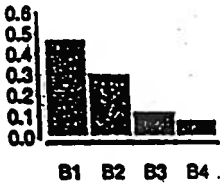
B3: Wendy's [.128]



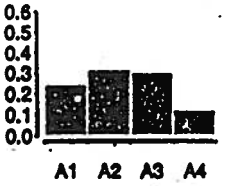
A4: 50+ [.126]



B4: Other [.088]



r.T[illegible]



c.Total [ ]

	B1	B2	B3	B4	c.Total
A1	0.15	0.068	0.02	0.012	0.25
A2	0.178	0.084	0.038	0.02	0.32
A3	0.108	0.104	0.056	0.036	0.304
A4	0.042	0.05	0.014	0.02	0.126
r.Total	0.478	0.306	0.128	0.088	1

P[i,j]

	B1	B2	B3	B4	c.Total
A1	0.1195	0.0765	0.032	0.022	
A2	0.1529	0.0979	0.0409	0.0281	
A3	0.1453	0.0930	0.0389	0.0267	
A4	0.0602	0.0385	0.0161	0.0110	
r.Total					

P.[i] x P.[j]

	B1	B2	B3	B4	c.Total
A1	75	34	10	6	125
A2	89	42	19	10	160
A3	54	52	28	18	152
A4	21	25	7	10	63
r.Total	239	153	64	44	500

Observed

Note if indep. Conditionals are the same

	B1	B2	B3	B4	c.Total
A1	0.6	0.272	0.08	0.048	1
A2	0.5562	0.2625	0.1187	0.0625	1
A3	0.3552	0.3421	0.1842	0.1184	1
A4	0.3333	0.3968	0.1111	0.1587	1
r.Total	0.478	0.306	0.128	0.088	1

Row.Conditionals

	B1	B2	B3	B4	c.Total
A1	0.3138	0.2222	0.1562	0.1363	0.25
A2	0.3723	0.2745	0.2968	0.2272	0.32
A3	0.2259	0.3398	0.4375	0.4090	0.304
A4	0.0878	0.1633	0.1093	0.2272	0.126
r.Total	1	1	1	1	1

Column.Conditionals

Joint

$$p(x,y) \equiv P(X=x \text{ and } Y=y)$$

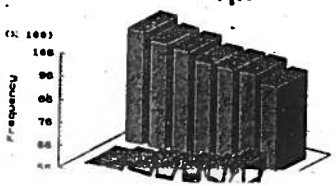
Marginal

$$p_X(x) \equiv P(X=x) = \sum_{all\ y} p(x,y)$$

Conditional

$$p_{Y|X}(y|x) \equiv P(Y=y \mid X=x) = \frac{p(x,y)}{p_X(x)}$$

$m(x) \equiv E(Y|X=x)$ and $v(x) \equiv Var[Y|X=x]$ are used to denote the mean and variance of the $p_{Y|X}(\cdot|x)$ conditional distribution.



GRS 3.4

Joint $f(x,y) \underset{\text{or}}{=} f_{X,Y}(x,y)$

Marginal $f_X(x) = \int_{\text{all } y} f(x,y)\,dy$

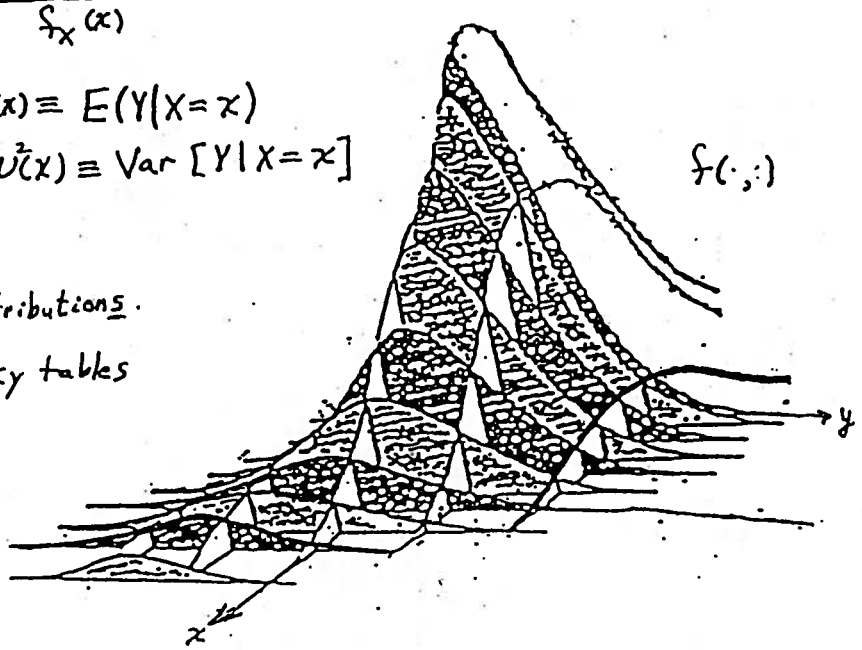
Conditional $f_{Y|X}(y|x) = \dfrac{f(x,y)}{f_X(x)}$

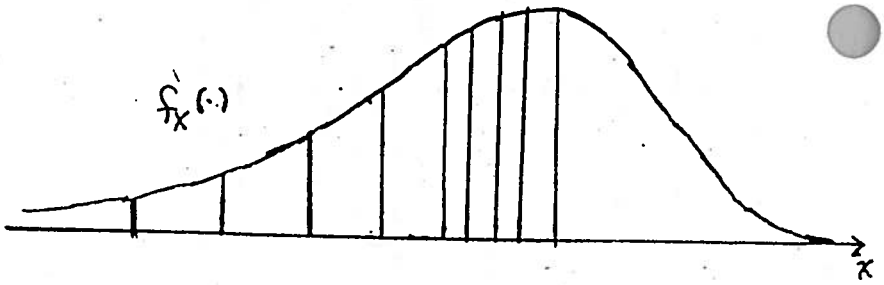
has mean $m(x) \equiv E(Y|X=x)$

and variance $v^2(x) \equiv \text{Var}[Y|X=x]$

This figure allows us to visualize the conditional distributions.

It is the analog of contingency tables in the discrete case.





Comment.

Probability. Given $f(\cdot,\cdot)$, compute things like $f_X(\cdot)$

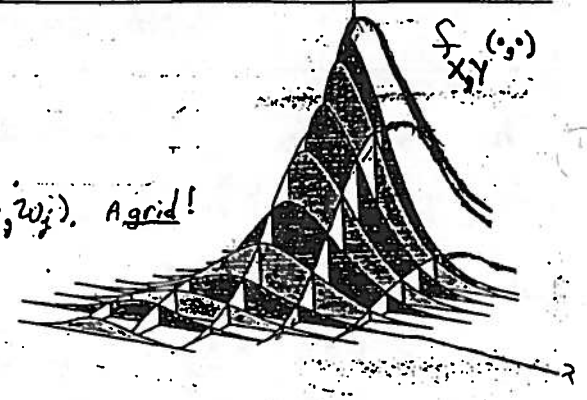
or $f_{Y|X}(\cdot|2)$ along with $m(2) = E(Y|X=2)$ and $v(2) = \text{Var}[Y|X=2]$.

Modeling

We determine an appropriate model for the $f(\cdot,\cdot)$ experiment via the two step procedure $f(x,y) = f_{Y|X}(y|x)\, f_X(x)$ where we use our scientific intuition (or our past work) to postulate the form of $f_X(\cdot)$ and $f_{Y|X}(\cdot|x)$ for the various x.

Joint and Marginal Densities

Suppose rv (X,Y) is given.



Discrete case.

The rv X has sample space $v_1, v_2, \ldots$.
The rv Y has sample space $w_1, w_2, \ldots$.
The sample space of (X,Y) is then the set of all (v_i, w_j). A grid!

$p_{X,Y}(v,w) \equiv P(X=v \text{ and } Y=w)$.

$P((X,Y)\in C) = \sum\sum_{\text{all } (v,w) \text{ in } C} p_{X,Y}(v,w)$

$p_X(v) \equiv P(X=v) = \sum_{\text{all } w} p_{X,Y}(v,w) \equiv$ (the marginal distribution of X).

$p_Y(w) \equiv P(Y=w) = \sum_{\text{all } v} p_{X,Y}(v,w) \equiv$ (the marginal distribution of Y).

Ex Red die X and White die Y. Consider $T=X+Y$ and $D=X-Y$.

X \ Y	Type of Crime: Homicide 1	Larceny 2	Assault 3	Burglary 4	
Area of City 1	30	120	450	150	750
2	25	200	800	300	1325
3	15	190	450	260	915
4	10	90	100	90	290
	80	600	1800	800	3280

Estimated $\hat{p}(v,w)$

X \ Y	1	2	3	4	
1	.0091	.0366	.1372	.0457	.2287
2	.0076	.0610	.2439	.0915	.4042
3	.0046	.0579	.1372	.0793	.2790
4	.0030	.0274	.0305	.0274	.0884
	.0244	.1829	.5488	.2439	

Continuous case.

The probability density $f_{X,Y}(v,w)$.

$P((X,Y)\in C) = \int_C\int f(x,y)\,dx\,dy$

$P(X\in A) = P(X\in A, Y\in(-\infty,\infty)) = \int_A\left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]dx = \int_A f_X(x)\,dx$.

$P(Y\in B) = P(X\in(-\infty,\infty), Y\in B) = \int_B\left[\int_{-\infty}^{\infty} f(x,y)\,dx\right]dy = \int_B f_Y(y)\,dy$.

So marginal densities are $f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$ and $f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx$.

Independent rv's X and Y: $P(X\in A \text{ and } Y\in B) = P(X\in A)P(Y\in B)$ for all A and B.

Thm. Discrete rv's independent if and only if $p_{X,Y}(v,w)=p_X(v)p_Y(w)$ for all v and w.

Continuous rv's independent if and only if $f_{X,Y}(x,y)=f_X(x)f_Y(y)$.

Pf for discrete: ($\Rightarrow$) $p_{X,Y}(v,w) = P(X\in\{v\} \text{ and } Y\in\{w\}) = P(X\in\{v\})P(Y\in\{w\}) = p_X(v)\,p_Y(w)$.
($\Leftarrow$) $P(X\in A \text{ and } Y\in B) \equiv \sum_{v\in A}\sum_{w\in B} p_{X,Y}(v,w) = \left\{\sum_{v\in A} p_X(v)\right\}\left\{\sum_{w\in B} p_Y(w)\right\} = P(X\in A)P(Y\in B)$. □

Def: $F_{X,Y}(x,y) \equiv P(X\le x \text{ and } Y\le y)$, with $F_{X,Y}(\cdot,\cdot)$ called the joint df.

Thm. Any rv's independent if and only if $F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$.

Fact: Also, for continuous rv's we have $f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{X,Y}(x,y)$.

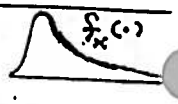
Convolution formula: If $T=X+Y$ where X and Y are independent rv's, then

$p_T(k) = \sum_{\text{all } i} p_X(i)\,p_Y(k-i)$ $\quad\left(=\sum_{i=0}^{k} p_X(i)\,p_Y(k-i) \text{ if } X\ge 0 \text{ and } Y\ge 0\right)$.

$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)\,f_Y(t-x)\,dx$ $\quad\left(=\int_0^t f_X(x)\,f_Y(t-x)\,dx \text{ if } X\ge 0 \text{ and } Y\ge 0\right)$.

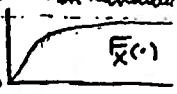
Continuous RV's.

One Dimension



The <u>density</u> $f_X(\cdot)$ gives a picture of where the probability is located.

The <u>df</u> $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ calculates all probability to the left of x. dt's used in calculus



We can <u>go</u> back from $F_X(\cdot)$ to $f_X(\cdot)$ by means of $f_X(x) = \frac{d}{dx} F_X(x)$.

Also, $P(X \in A) = \int_A f_X(x)\,dx$ allows us to calculate a probability.

Two Dimensions

Likewise $f_{X,Y}(\cdot,\cdot)$ gives a picture of where the <u>joint probability</u> is located.

Also, $P((X,Y) \in A) = \int_A \int f_{X,Y}(x,y)\,dx\,dy$ allows to calculate the probability of an event A.

<u>Marginals</u>: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$.

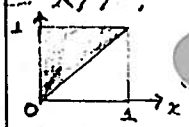
<u>Conditionals</u>: $f_{Y|X=x}(y) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$ for all y and $f_{X|Y=y}(x) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$ for all x.

We can also use these conditionals to tell us about the joints (just cross multiply)

$f_{X,Y}(x,y) = f_X(x)\, f_{Y|X=x}(y)$ or $f_{X,Y}(x,y) = f_Y(y)\, f_{X|Y=y}(x)$. (math models)

Maybe we want to consider some <u>transformation</u> of X and Y like $V \equiv X/Y$.

<u>Example 4.</u> We're given $f_{X,Y}(x,y) = 1/y$ for $0 \le x \le y \le 1$.



Then $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^y \frac{1}{y}\,dx = \frac{x}{y}\Big|_0^y = 1$ for all $0 < y \le 1$. Thus $Y \cong \text{Uniform}(0,1)$.

Also $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_x^1 \frac{1}{y}\,dy = \log y\Big|_x^1 = 0 - \log x = \log \frac{1}{x}$ for all $0 < x \le 1$. This density is new to us.

For fun, $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x (-\log t)\,dt = (-t\log t + t)\Big|_0^x = x(1 - \log x)$ for $0 < x \le 1$.

How do we find the probability of an event like $A \equiv [Y > 2X]$?

$$P(Y>2X) = \iint_{y>2x} f(x,y)\,dx\,dy = \int_0^1 \int_0^{y/2} \frac{1}{y}\,dx\,dy = \int_0^1 \frac{1}{2}\,dy = \frac{1}{2}.$$

or

$$P(Y>2X) = \text{ditto} = \int_0^{1/2} \int_{2x}^1 \frac{1}{y}\,dy\,dx = \int_0^{1/2} (-\log 2x)\,dx = (-x\log 2x + x)\Big|_0^{1/2} = 1/2.$$

Let's transform to the variable $V \equiv X/Y$. Then

$$F_V(v) = P(V \le v) = P(X/Y \le v) = \iint_{x \le vy} f(x,y)\,dx\,dy = \int_0^1 \int_0^{vy} \frac{1}{y}\,dx\,dy = \int_0^1 v\,dy = v \quad \text{for } 0 < v < 1$$

and so

$f_V(v) = F_V'(v) = \frac{d}{dv} v = 1$ for $0 < v < 1$. Thus $V \cong \text{Uniform}(0,1)$.

Note: $f_{X,Y}(x,y) = f_{X|Y=y}(x) = f_Y(y) = \frac{1}{y} f_U \quad 0 \le x \le y$

$= \left\{ \frac{1}{y} 1_{[0,y]}(x) \right\} \cdot 1$ uniform $(0,1)$

$= f_{X|Y=y}(x) \cdot f_Y(y) = f_{u(0,y)}(x) \cdot f_{u(0,1)}(y)$

$=$

Conditional Expectation

Given the joint density $f(x,y)$ of rvs X, Y.

The conditional density is

$$f_{Y|X=x}(y) \equiv f(y|x) \equiv \frac{f(x,y)}{f_X(x)} \qquad \text{for "all" } y, \quad \text{for "each" } x;$$

and so

$$f(x,y) = f_{Y|X=x}(y)\, f_X(x) \qquad \text{for "all" } x, y.$$

Remember, the conditional density is a density; we can thus talk about its mean and variance, etc. Recall first that

μ_X, μ_Y and σ_X^2, σ_Y^2 denote the means and variances of f_X and f_Y.

We now agree that:

conditional mean: $m(x) \equiv E(Y|X=x) = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\,dy$ denotes the mean of $f_{Y|X=x}(\cdot) = f(\cdot|x)$.

$m(\cdot) \equiv E(Y|X=\cdot) \equiv E(Y|X)$.

conditional variance: $v^2(x) \equiv \mathrm{Var}[Y|X=x]$ denotes the variances of $f_{Y|X=x}(\cdot) = f(\cdot|x)$.

$v^2(\cdot) \equiv \mathrm{Var}[Y|X=\cdot] = \mathrm{Var}[Y|X]$.

$E(g(Y)|X=x) \equiv E(g(Y)\,|\,X=x) \equiv E(g(Y)|x)$ denotes the mean of $g(Y)$ wrt $f_{Y|X=x}(\cdot)$.

Theorem

(1) $\mu_Y = EY = E(E(Y|X)) = Em(X) = \begin{cases} \int_{-\infty}^{\infty} m(x) f_X(x)\,dx & \text{if } X \text{ is continuous,} \\ \sum_{\text{all } x} m(x) p_X(x) & \text{if } X \text{ is discrete.} \end{cases}$

(2) $\sigma_Y^2 = \mathrm{Var}[Y] = E(\mathrm{Var}[Y|X]) + \mathrm{Var}[E(Y|X)] = Ev^2(X) + \mathrm{Var}[m(X)]$

Corollary $E(Z) = E(E(Z|X))$, or (for example) $EY^2 = E(E(Y^2|X))$.

$\mathrm{Var}[Y] = E(E(Y^2|X)) - \mu_X^2$. ← often this is sufficient (let $Z = Y^2$)

Proof. In the continuous case,

$$\mu_Y = \int y f_Y(y)\,dy = \int y \int f(x,y)\,dx\,dy = \int \left[\int y \frac{f(x,y)}{f_X(x)}\,dy\right] f_X(x)\,dx$$
$$= \int \left[\int y\, f_{Y|X=x}(y)\,dy\right] f_X(x)\,dx = \int m(x) f_X(x)\,dx = Em(X) = E(E(Y|X)).$$

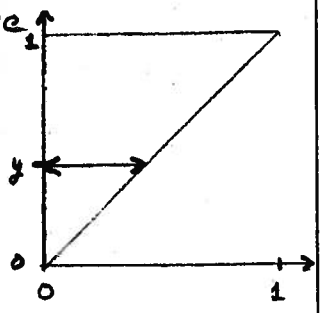
In general,

$$\sigma_Y^2 = EY^2 - \mu_Y^2 = E(E(Y^2|X)) - \mu_Y^2 = E\left(E(Y^2|X)\right) - E^2 m(X)$$
$$= E\{\underbrace{E(Y^2|X) - m^2(X)}_{\text{the variance of } P_{Y|X}}\} + \underbrace{Em^2(X) - E^2 m(X)}_{\text{the variance of } m(X)}$$
$$= Ev^2(X) + \mathrm{Var}[m(X)].$$

□

Example 1. Let

$$f(x,y) = \frac{1}{y} 1_{[0\le x\le y\le 1]} = \begin{cases} \frac{1}{y} & \text{for } 0\le x\le y\le 1 \\ 0 & \text{else} \end{cases}$$



$$= \frac{1}{y}\cdot 1_{[0\le x\le y]} \times 1_{[0,1]}(y)$$

$$= f_{\text{Uniform}(0,y)}(x) \times f_{\text{Uniform}(0,1)}(y)$$

$$= f_{X|Y=y}(x) \times f_Y(y)$$

mean $= y/2 = m(y)$, variance $= y^2/12 = v^2(y)$ | mean $= 1/2$, variance $= 1/12$.

Thus

$$\mu_X = Em(Y) = E(\tfrac{Y}{2}) = \tfrac{1}{4}.$$

$$\sigma_X^2 = Ev^2(Y) + \text{Var}[m(Y)] = E\tfrac{Y^2}{12} + \text{Var}[\tfrac{Y}{2}] = \tfrac{1}{12}\left(\tfrac{1}{12}+(\tfrac{1}{2})^2\right) + \tfrac{1}{4}\cdot\tfrac{1}{12} = \tfrac{7}{144}.$$

$$\text{Cov}[X,Y] = E(XY) - \mu_X\mu_Y = E\big(E(XY|Y)\big) - \tfrac{1}{4}\cdot\tfrac{1}{2} = E\big(YE(X|Y)\big) - \tfrac{1}{8}$$

$$= E(Y\cdot\tfrac{Y}{2}) - \tfrac{1}{8} = \tfrac{1}{2}\left(\tfrac{1}{12}+(\tfrac{1}{2})^2\right) - \tfrac{1}{8} = \tfrac{1}{6}\cdot\tfrac{1}{8} = \tfrac{1}{24}. \qquad \square$$

Example 2. Let

$$f(x,y) = xe^{-x(y+1)} \quad \text{for } x\ge 0 \text{ and } y\ge 0.$$

$$= xe^{-xy}1_{[y\ge 0]} \times e^{-x}1_{[x\ge 0]}$$

(Poisson)

$$= f_{\text{Exp}(\text{rate}=x)}(y) \times f_{\text{Exp}(1)}(x)$$

$$= f_{Y|X=x}(y) \times f_X(x)$$

mean $= \frac{1}{x} = m(x)$, variance $= \frac{1}{x^2} = v^2(x)$ | mean $= 1$, variance $= 1$.

Thus

$$EY = Em(X) = E(\tfrac{1}{X}) = \int_0^\infty \tfrac{1}{x}e^{-x}dx \ge \int_0^1 \tfrac{e}{x}\,dx \ge e\log x\Big|_0^1 = \infty.$$

Or, $EY = \int_0^\infty\int_0^\infty y\,xe^{-x(y+1)}\,dy\,dx = \int_0^\infty\left[\int_0^\infty \tfrac{x^2y}{\Gamma(2)}e^{-xy}dy\right]\tfrac{1}{x}e^{-x}dx = \int_0^\infty \tfrac{1}{x}e^{-x}dx = \infty.$

What is the density of XY?

Instead, what is the conditional density of $XY|X=x$?

Or, what is the conditional density of xY given $X=x$?

Answer: xY is Exponential (1), when x is regarded as fixed.

Since xY is $\text{Exp}(1)$ for every x, $XY \cong \text{Exp}(1)$ also.

Pf. $P(XY\le t) = E(1_{[XY\le t]}) = E\big(P(XY\le t|X)\big) = E\big(P(\text{Exp}(1)\le t)\big) = P(\text{Exp}(1)\le t)$.

So, $XY \cong \text{Exp}(1)$.

Comment. Only one factorization $f(x,y) = f_{Y|X=x}(y)\,f_X(x)$ is possible.

Did HW #1 (from L14)

Suppose X and Y have joint density

$$f(x,y) = xe^{-x(y+1)} \quad \text{for } x>0 \text{ and } y>0.$$

(a) Now

$$f(x,y) = xe^{-xy}\cdot e^{-x}$$

$$= f_{Y|X=x}(y)\cdot f_X(x) \qquad \left\{= g(x) f_{Y|X=x}(y)\cdot \frac{f_X(x)}{g(x)} \right\}$$

(*) $\cong$ Exp(rate $=\nu=x$) $\cdot$ Exp(1) — can not happen.

(That is, if you find one factorization that works, then it is the only one.)

Now the conditional density $f_{Y|X=x}(\cdot)$ has:

mean $= m(x) = E(Y|X=x) = \{$mean of Exp(rate $=\nu=x)\} = \frac{1}{x}$.

variance $= v^2(x) = \mathrm{Var}[Y|X=x] = \{$variance of Exp(rate $=\nu=x)\} = \frac{1}{x^2}$.

Thus

$$EY = E(m(X)) = \int_0^\infty m(x) f_X(x)\,dx = \int_0^\infty \frac{1}{x} e^{-x}\,dx$$

$$= \int_0^1 \frac{1}{x} e^{-x}\,dx + \int_1^\infty \frac{1}{x} e^{-x}\,dx \geq \frac{1}{e}\underbrace{\int_0^1 \frac{1}{x}\,dx}_{=\log x\,|_0^1 = 0-(-\infty)=\infty} + 0 = \infty.$$

Note that:

$$EY = \int_0^\infty\int_0^\infty y\, xe^{-x(y+1)}\,dy\,dx = \int_0^\infty \left[\int_0^\infty y\, xe^{-xy}\,dy\right] e^{-x}\,dx$$

(Gamma) $$= \int_0^\infty \frac{1}{x}\underbrace{\left[\int_0^\infty \frac{x^2 y}{\Gamma(2)} e^{-xy}\,dy\right]}_{=1,\text{ the hard way (via Gamma}(\nu=x, r=2))} e^{-x}\,dx = \int_0^\infty \frac{1}{x} e^{-x}\,dx = \infty.$$

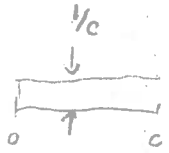
(b) What is the density of XY?

Instead, we ask: What is the conditional density of $XY\,|\,X=x$? That is, what is the density of xY, for each x?

Answer: Since xY is Exp(1) (note (*)) for every x, the density of XY is Exp(1) also.

Uniform Distribution

$\mu = \frac{c}{2}$ $\sigma^2 = \frac{c^2}{12}$



$E(XY|Y=y) = yE(X|Y=y)$

find a to minimize $E(Y-a)^2 = E\left((Y-\mu)+(\mu-a)\right)^2$

HW#1 (cont) $f(x,y) = xe^{-x(y+1)}$ ← (looks a bit like a gamma)

$$= \frac{(y+1)^2 x e^{-(y+1)x}}{\Gamma(2)} \cdot \underbrace{\frac{1}{(y+1)^2}}_{\text{Quick check! Yes!! It is a density.}} \quad \Gamma(2) = 1$$

$$= f_{X|Y=y}(x) \cdot f_Y(y)$$

$$\cong \text{Gamma}(\nu = y+1, r=2) \cdot (\text{No name})$$

But $m(y) \equiv E(X|Y=y) = \{\text{mean of Gamma}(\nu = y+1, r=2)\}$

$$= \frac{r}{\nu} = \frac{2}{y+1}.$$

Thus we can claim that:

$$EX = Em(Y) = E\left(\frac{2}{Y+1}\right) = \int_0^\infty \frac{2}{y+1} f_Y(y)\,dy = \int_0^\infty \frac{2}{y+1} \cdot \frac{1}{(y+1)^2}\,dy$$

$$= \int_0^\infty \frac{2}{(y+1)^3}\,dy = -\frac{1}{(y+1)^2}\Big|_0^\infty = -0 - -1 = 1;$$

but we already knew this (since $X \cong Exp(1)$).

Homework

#1. (a) Calculate $v^2(y) \equiv Var[X|Y=y]$.

(b) Then calculate $Var[X] = Var[m(Y)] + E[v^2(Y)]$.

[You know the answer has to be 1, since $X \cong Exp(1)$.]

(c) We showed above that $f_Y(y) = \frac{2}{(y+1)^2}$ for $y \geq 0$.

Calculate EY.

[You know the answer has to be ∞ (by (**)).]

Conditional Probability

$$P(A) = E(1_A) = E\{E(1_A|X)\} = \int_{-\infty}^{\infty} E(1_A|X=x) f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} P(A|X=x) f_X(x)\,dx = E(P(A|X)).$$

For problems that can be one or zero

Homework from Casella and Berger

Chapter 4 (#4.12 and #4.13).

#1. (10) What is the probability that the sum of two dice (one is red and one is white) equals a prime number, given that the red one is odd?

#2. (20) Let ν denote the rate at which craters visible to our telescope are created in the Sea of Tranquility (on the moon), measured in time units of one year.
(a) Suppose $\nu = 10$.
Evaluate the probability of seeing at least 12 new craters during the next 2 years.
(b) Given that exactly 8 occur in each of the next 2 years, what is the probability that 2 or less occur in the first 1/4 of each of the two years.

#3. (20) Consider a Poisson process with intensity ν.
Let $Y_1, Y_2, \ldots$ denote the interarrival times.
Suppose that $T \equiv Y_1 + \cdots + Y_r$ for some integer r.
(a) What is the distribution of T?
(b) Determine $\mathrm{E}\left(\frac{1}{T}\right)$, (by showing your calculations).

#4. (35) Let X have the density $f(x) = \alpha/(1+x)^{1+\alpha}$ for $x \geq 0$, for some value of $\alpha > 0$.
(a) Determine the density of the rv $Y \equiv \log(1+X)$.
(b) Evaluate $F(x)$ for each $x \geq 0$.
(c) Evaluate $F^{-1}(t)$ for each $0 < t < 1$.
(d) Suppose that X and Y are independent and both have the density $f(\cdot)$.
Evaluate $P(X < Y)$.
(e) Suppose that X and Y are independent and both have the density $f(\cdot)$.
Evaluate $P(\min(X,Y) \leq x)$ for $x \geq 0$.

#5. (15) Let X denote a rv with mgf $M(t)$ that is finite in a neighborhood of the origin.
Define $\psi(t) \equiv \log M(t)$.
(a) What are the values of $M(0)$ and $M'(0)$?
(b) What is the value of $\psi'(t)\Big|_{t=0}$?
[You may express it in terms of parameters or moments of the distribution of X.]
(c) What is the value of $\psi''(t)\Big|_{t=0}$?
[You may express it in terms of parameters or moments of the distribution of X.]

#6 A deck of "Rook cards" contains 14 cards (numbered $1, 2, \ldots, 14$) in each of the four colors Black, Red, Green, Yellow; thus, there are $14 \times 4 = 56$ cards in this deck. You are given a "hand" (that is, a subset) consisting of size 10 cards. Let

(1) $X \equiv$ (the number of Black cards in you hand of size 10)
$Y \equiv$ (the number of Red cards in you hand of size 10)

(a) (10) Define the joint mass function

(2) $f_{X,Y}(x,y) \equiv P(X = x \text{ and } Y = y)$ for appropriate x, y.

.....

In parts (b) and (c) below, you can revisualize so as to answer all questions with essentially no calculations.

.....

(b) (9) State the name of the conditional distribution of X given that $Y = y$?
Then write out an expression for the probability

(3) $f_{X|Y=y}(x) \equiv P(X = x | Y = y)$.

(c) (6) State a formula for the mean of this conditional distribution. (No need to derive it.)

#7. (25) Telephone calls arrive at our switchboard at the rate of 0.5 per minute. There are 30 two minute periods per hour (hour after hour), ... In each of these two minute periods we say that success occurs if the number of calls that arrives is equal to zero. (Time starts at 12:00.)
(a) Evaluate numerically the probability p of success in the interval from 12:00 to 12:02.
(b) Evaluate the probability that the first success occurs in the time period from 12:06 to 12:08.
(c) Evaluate the probability of no successes by 12:08.
(d) Suppose that by 1:00 there have been 20 successes. Let U denote the number of successes by 12:30. Name the distribution of U.

#8. Suppose X and Y have joint density

$$f(x,y) = 2e^{-2x} \frac{y^3}{3!\,x^4} e^{-y/x} \quad \text{for } y>0 \text{ and } x>0$$

(a) Give the name and formula for $f_X(\cdot)$. Then $\mu_X =$ _____, and $\sigma_X^2 =$ _____.

(b) Give the name and formula for $f_{Y|X=x}(\cdot)$.

(c) $m(x) \equiv E(Y|X=x) =$ _____. And, $EY =$ _____.

$v^2(x) \equiv \mathrm{Var}[Y|X=x) =$ _____.

(d) Now that you know a formula for $m(X)$,

$\mathrm{Var}[m(X)] =$ _____.

(e) Now that you know a formula for $v^2(X)$,

$E\{v^2(X)\} =$ _____.

(f) So, $\mathrm{Var}[Y] =$ _____.

#9. Suppose X and Y have joint density

(*) $$f_{X,Y}(x,y) = \frac{y^2}{2} e^{-y} \frac{2x}{y^2} \quad \text{on } 0 \le x \le y < \infty.$$

(a) Determine $f_Y(y) =$ _____ on _____, and $f_{X|Y}(x|y) =$ _____ on _____.

(b) Determine $E(X|Y=y) =$ _____ and $\mathrm{Var}[X|Y=y] =$ _____.

Define $U = X/Y$ and $V = Y$ for the X, Y of (*).

(a) Determine that $f_{U,V}(u,v) =$ _____ on _____.

(b) Determine $f_{U|V}(u|v) =$ _____ on _____.

"PLANNED" Syllabus for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29	L1...Ch1: Probability, conditional probability, independence, and rvs
Mar 31	L2...Sec2.1–2.3$^-$: E(·), mean, variance, and (especially) transformations (no mgf)
Apr 01	Q....HW and some catchup lecturing
	L3...Sampling with and without replacement
.....	
Apr 05	L4...Sec3.1–3.2(Handout): Bernoulli trials; Geometric, NegBinomialT (and Hypergeometric and Discrete Uniform)
	L5...Sec3.3(Handout): Poisson process, Poisson, Exponential, Gamma (and Beta and Uniform)
	Q ....HW, and finish Poisson process
	L6...Sec2.3: Moment generating function (mgf)
.....	
Apr 12	L7...More mgf, and CLT pics
	L8...Ch4: Joint, marginal, and conditional distributions
	Q....Sec4.1: Conditionals and convolution
	L9...Sec4.2: Conditional mean and variance
.....	
Apr 19	L10..Sec4.2 Independence and Expectation
	L11..Other Ch4 material
	Q....MIDTERM I
	L12..Discuss Midtem, and begin anew
.....	
Apr 26	L13..Sec4.5: Covariance
	L14..Sec5.1–5.2: Properties of $\bar{X}$ and S^2
	Q....
Apr 30	L15..Sec4.5: Bivariate Normal Distribution
.....	
May 03	L16..Sec5.3: Derivation of Student-T and Fisher/Snedecor/varianceRatio-F
	L17..Handout: Moments via matrices
	Q....
	L18..Sec7.1–7.2:Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)
.....	
May 10	L19..MLE: Three illustrative MLE examples
	L20..Sec6.1–6.2: Sufficiency
	Q....Work HW on MLE
	L21..Completenes, Exponential families, ancillarity
.....	
May 17	L22..Bayes estimators
	L23..Bayes examples
	Q....MIDTERM II
	L24..Discuss Midterm, and begin anew
.....	
May 24	L25..Cramer–Rao inequality
	L26..Finish Cramer–Rao
	Q....
	L27..Hypothesis tests
.....	
May 31	MEMORIAL DAY HOLIDAY
June02	L28..Likelihood ratio tests
	Q....Work examples of LR-tests
June04	L29..Location and scale models
.....	
June09	FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....	

Planned Homework for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Chapter1: #[2,5,11,13,14,25,33,38; do not be formal; should be easy];#17,18,21,23,24
Mar 31 L2...Chapter1: #44,53; Chapter2: #4,8c,17,19,24
Apr 01 Q....
L3...Chapter2: #20; Chapter3: #4,6,12,23,24b; Handout: #1
.....
Apr 05 L4...Handout: #2–5
L5...Handout: #6–7
Q ...
L6...
.....
Apr 12 L7...Chapter2: #9,16, (read 11); Chapter3: #22cd
L8...Chapter2: #33abc; Chapter3: #19,22ab
Q....
L9...Chapter2: # 26ab, 28ab, 30ad; Chapter3: #24c
.....
Apr 19 L10..Chapter4 #13,15; HW #1abc in Conditional Epectation handout about $v^2(y)$
L11..Unofficial: Work some off the HW sheet containing nine
Q....MIDTERM
L12..
.....
Apr 26 L13..
L14..
Q....
Apr 30 L15..
.....
May 03 L16..
L17..
Q....
L18..
.....
May 10 L19
L20
Q....
L21..
.....
May 17 L22..
L23..
Q....
L24..
.....
May 24 L25..
L26..
Q....
L27..
.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28..
Q
June04 L29..
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

EXAMPLE

6 R	
3 W	X
5 B	Y

$n=6$

Compute $f_{X|Y=3}(x)$

6R
3W
$n=3$

Hypergeometric(6R, 3W; $n=3$)

so $f_{X|Y=3}(k) = \dfrac{\binom{3}{k}\binom{6}{3-k}}{\binom{9}{3}}$

0	1	2	3
$\frac{20}{84}$	$\frac{45}{84}$	$\frac{18}{84}$	$\frac{1}{84}$

$m(3) = E(X|Y=3) = np = 3\cdot\frac{1}{3} = 1$

$v^2(3) = \mathrm{Var}[X|Y=3] = npq\left[1-\frac{n-1}{N-1}\right] = 3\cdot\frac{1}{3}\cdot\frac{2}{3}\left[1-\frac{2}{8}\right] = \frac{1}{2}$

Compute $f_{X|Y=1}(x)$

6R
3W
$n=5$

Hypergeometric (6R, 3W; $n=5$)

$m(1) = E(X|Y=1) = np = 5\cdot\frac{1}{3} = \frac{5}{3}$

$v^2(1) = \mathrm{Var}[X|Y=1] = npq\left[1-\frac{n-1}{N-1}\right] = 5\cdot\frac{1}{3}\cdot\frac{2}{3}\left[1-\frac{4}{8}\right] = \frac{5}{9}$. □

EXAMPLE Suppose $f(x,y) = 4y(x-y)e^{-(x+y)}$ on $0 \le y \le x < \infty$

Tell me about it?

Soln. Well

$$f(x,y) = 4y(x-y)e^{-(x-y)}e^{-2y}$$
$$= 2^2 y e^{-2y} \cdot (x-y)e^{-(x-y)}$$
$$= f_Y(y) \cdot f_{X|Y=y}(x)$$

$Y \cong \text{Gamma}(r=2, \nu=2)$ $\qquad$ $X-y \cong \text{Gamma}(r=2, \nu=1)$

$\mu_Y = \frac{r}{\nu} = \frac{2}{2} = 1$ $\qquad$ $m(y) - y = \frac{r}{\nu} = 2$ or $m(y) = y+2$

$\sigma_Y^2 = \frac{r}{\nu^2} = \frac{2}{4} = \frac{1}{2}$ $\qquad$ $v^2(y) = \frac{r}{\nu^2} = 2$

So $\mu_X = E\, m(Y) = E(Y+2) = 1+2 = 3.$

$\sigma_X^2 = E(v^2(Y)) + \text{Var}[m(Y)] = E(2) + \text{Var}[Y+2] = 2 + \frac{1}{2} = \frac{5}{2}.$ $\square$

INDEPENDENCE

The following are equivalent:

(1) X and Y are independent rv's {words in search of a definition}.

(2) $P(X\in A \text{ and } Y\in B) = P(X\in A)P(Y\in B)$ for all 1-dim events A and B.

(3) $P(X\le x \text{ and } Y\le y) = P(X\in x)P(Y\le y)$ for all real x,y [$A=(-\infty,x]$, $B=(-\infty,y]$]

$F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$ for all real x,y.

(4) $f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$ for all real x,y; if ....

(5) $\phi_{X,Y}(s,t) = E\{e^{i(sX+tY)}\} = \phi_X(s)\,\phi_Y(t)$ for all real s,t.

{(6) $M_{X,Y}(s,t) \equiv E\{e^{sX+tY}\} = M_X(s)\,M_Y(t)$ for all reals s,t.}; if these exist...

Note that:
(1) is an idea.
(2) is the natural definition.
(3)-(6) are theorems.
Or use any other "one" for a definition, and the rest for theorems.
This is also true if $X=(X_1,\dots,X_m)$, $Y=(Y_1,\dots,Y_n)$, $s=(s_1,\dots,s_m)$, $t=(t_1,\dots,t_n)$

"Proof". (2) $\Rightarrow$ (3). This is trivial.

(3) $\Rightarrow$ (2). By HW 4.9 we have (3) implies $P(a<X\le b \text{ and } c<Y\le d) = P(a<X\le b)P(c<Y\le d)$

Thus $P(X\in A, Y\in B) = P(X\in A)P(Y\in B)$ for all $A=(a,b]$ and $B=(c,d]$.

Thus $P\left((X,Y)\in \sum_1^m (A_i\times B_j)\right) = \sum_1^m P(X\in A_i)P(Y\in B_j)$ for all disjoint $A_i\times B_i = (a_i,b_i]\times(c_i,d_i]$

Thus $P((X,Y)\in A\times B) = P(X\in A)P(Y\in B)$ for all A event and all B events. (See σ-field page 6)

(4) $\Rightarrow$ (3) This is trivial. (if $F_{X,Y}$ has a density or a mass function).

Continuous case:

i.e. $P(X\le x \text{ and } Y\le y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(u,v)\,d(u\times v) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_X(u)\,f_Y(v)\,d(u\times v)$

$= \int_{-\infty}^{x} f_X(u)\,du \int_{-\infty}^{y} f_Y(v)\,dv = F_X(u)\,F_Y(v)$.

Discrete case: $\int dx \to \sum$.

(3) $\Rightarrow$ (4) (The idea!)

Continuous case: $f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_X(x)\,F_Y(y)$

$= \frac{\partial}{\partial x}F_X(x)\,\frac{\partial}{\partial y}F_Y(y) = f_X(x)\,f_Y(y)$.

(3) $\Leftrightarrow$ (2) $\Rightarrow$ (4)

Discrete case: $f_{X,Y}(x,y) = P(X=x \text{ and } Y=y) = P(X=x)P(Y=y) = f_X(x)\,f_Y(y)$

(2) $\Rightarrow$ (5) is trivial. (5) $\Rightarrow$ (2) comes out of the inversion formula.
{Add (6) to the list in the same way.}

Independence

Events A and B were called *independent* if $P(AB) = P(A)P(B)$. We extend this idea to rv's.

Theorem 1. X and Y are *independent rv's* if and only if

(def) ... $P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$ ... for all A and B ...

(a) $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all x,y OR $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x,y.

(b) $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all x,y.

Example 5. ($X \equiv$ (outcome from a die) and $Y \equiv$ (outcome from a coin))

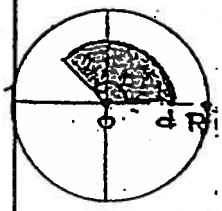
	1	2	3	4	5	6	
H=1							$\frac{1}{2}$
T=0							$\frac{1}{2}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1.00

Fair coin and fair die

	1	2	3	4	5	6	
H=1							$\frac{1}{3}$
T=0							$\frac{2}{3}$
	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	

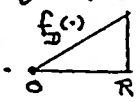
Biased coin and biased die

Example 6. (Choose a point at random from the circle.)

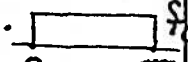


Let $D \equiv$ (distance from the origin) and $\Theta \equiv$ (angle from the origin)

Now $F_D(d) = P(D \le d) = \frac{\pi d^2}{\pi R^2} = d^2/R^2$, so $f_D(d) = 2d/R^2$ for $0 \le d \le R$.



Also $F_\Theta(t) = P(\Theta \le t) = \frac{t}{2\pi}$ for $0 \le t \le 2\pi$, so $f_\Theta(t) = \frac{1}{2\pi}$ for $0 \le t \le 2\pi$.



Also $F_{D,\Theta}(d,t) = P(D \le d \text{ and } \Theta \le t) = \frac{\pi d^2}{\pi R^2} \cdot \frac{t}{2\pi} = \frac{t d^2}{2\pi R^2}$, $= F_\Theta(t)F_D(d)$ for independence.

Thus $f_{D,\Theta}(d,t) = \frac{\partial}{\partial t}\frac{\partial}{\partial d} F_{D,\Theta}(d,t) = \frac{\partial}{\partial t}\frac{t 2d}{2\pi R^2} = \frac{1}{2\pi}\frac{2d}{R^2}$, $= f_\Theta(t)f_D(d)$ for independence.

Example 9 (Buffon's needle)

Convolution Formula. Let $Z \equiv X + Y$, for independent X and Y.

$$p_Z(z) = \sum_{\text{all } x} p_X(x)\,p_Y(z-x) \quad \{\text{or} = \sum_{\text{all } y} p_Y(y)\,p_X(z-y)\} \quad \text{(discrete)}$$

$$f_Z(z) = \int_{\text{all } x} f_X(x)\,f_Y(z-x)\,dx \quad \{\text{or} = \int_{\text{all } y} f_Y(y)\,f_X(z-y)\,dx\} \quad \text{(continuous)}$$

EXPONENTIAL FAMILY

GRS

Definition 1. (Exponential family) Suppose we have densities $f_\theta(\cdot)$, $\theta \in \Theta$, of the form

$$f_\theta(x) = f(x|\theta) = C(\theta)\, e^{\sum_{i=1}^{k} w_i(\theta)\, T_i(x)}\, h(x)$$

for function $T_1, \ldots, T_k$ of the observations, and functions $C(\theta)$, $w_1(\cdot), \ldots, w_k(\cdot)$ of the parameters. Then $\{f_\theta : \theta \in \Theta\}$ is a <u>exponential family</u>.

Definition 2 (Natural parameter) For $\eta_1(\theta), \ldots, \eta_k(\theta)$, we may reparameterize:

$$= C^*(\eta)\, e^{\sum_{i=1}^{k} \eta_i\, T_i(x)}\, h(x).$$

Note: The <u>natural parameter space</u> (values of η for which this is indeed a density) is <u>convex</u>, and $E_\eta\{g(x)\}$ is an <u>analytic function</u> on the interior.

Example 1 (Bernoulli trials, $0 < p < 1$)

$$f_p(x) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}$$

$$= (1-p)^n\, e^{(\sum_1^n x_i)\log\frac{p}{1-p}} \cdot 1$$

$$= C^*(\eta)\, e^{T(x)\,\eta} \cdot 1$$

where $\eta \equiv \eta(p) \equiv \log\frac{p}{1-p} \equiv$ (<u>log odds</u>). (Note: $p = \frac{1}{2}$ iff $\eta = 0$.)

[Comment: Binomial(n,p) has $h(x) = \binom{n}{x}$.]

Comment. $\underline{X} \equiv (X_1, \ldots, X_n)'$ are iid Bernoulli(p); so $T \equiv T(\underline{X}) \equiv$ Binomial(n,p).

Then $f_{\underline{X}|T=t}(\underline{x}) = \dfrac{p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}}{\binom{n}{t} p^t (1-p)^{n-t}}\, 1_{[\sum_1^n x_i = t]} = \dfrac{1}{\binom{n}{t}}\, 1_{[\sum_1^n x_i = t]}$

so that this conditional distribution is independent (is not a function of) θ. That is, T is <u>sufficient</u> for θ.

The dist of the suff stat is also an exponential family

Example 2. (The $N(\mu, \sigma^2)$ family) Now

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}}\right] e^{-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}} \cdot 1$$

$$= c^*(\eta)\, e^{\eta_1 x^2 + \eta_2 x} \cdot 1 = c^*(\eta)\, e^{\eta_1 T_1(x) + \eta_2 T_2(x)}.$$

Now let $X_1, \ldots X_n$ be iid $N(\mu, \sigma^2)$. Then

$$f_{\mu,\sigma^2}(\underline{x}) = \ldots = \left[\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{n\mu^2}{2\sigma^2}}\right] e^{-\frac{\sum_1^n x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum_1^n x_i},$$

in the same spirit as before. [Note: $\mu = 0$ iff $\eta_2 = 0$ (too bad not $\frac{\mu}{\sigma^2}$).

Shorack STAT 506 June 9, 2004

Final Exam (closed book and closed notes).

Suppose that a basix X-experiment has density

$$f_\theta(x) = \frac{\theta}{(1+x)^{1+\theta}} \qquad \text{for } x \geq 0 \qquad \text{(for each parameter value } \theta > 0\text{)}.$$

This *is* a "regular family" (that is, all the math works). [Moreover, the math is easy.]
(It could be helpful to make the "abbreviation" that $Y \equiv \log(1+X)$.)

#1. (100)
(a) Compute the score function $\dot{\ell}_\theta \equiv \dot{\ell}(\theta; X)$ for one observation X.
What is the mean of the score function always equal to (in any regular model)?
(What do you learn from this latter fact?)
(b) Calculate the information $\bar{\mathcal{I}}(\theta)$ (for one observation).
(c) Tell me the *value* of $E_\theta\{[\dot{\ell}(\theta; X)]^2\}$.
(There is no need to do any additional calculation here. Tell me why.)
(d) You have not yet determined the MLE $\hat{\theta} = \hat{\theta}_n$. Nevertheless, fill in the following blank:

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \to_d \ldots\ldots\ldots\

EXPONENTIAL FAMILY

GRS

Definition 1. (Exponential family) Suppose we have densities $f_\theta(\cdot)$, $\theta \in \Theta$, of the form

$$f_\theta(x) = f(x|\theta) = C(\theta)\, e^{\sum_{i=1}^{k} w_i(\theta) T_i(x)}\, h(x)$$

for function $T_1, \ldots, T_k$ of the observations, and functions $C(\theta)$, $w_1(\cdot), \ldots, w_k(\cdot)$ of the parameters. Then $\{f_\theta : \theta \in \Theta\}$ is a exponential family.

Definition 2 (Natural parameter) For $\eta_1(\theta), \ldots, \eta_k(\theta)$, we may reparameterize:

$$= C^*(\eta)\, e^{\sum_{i=1}^{k} \eta_i T_i(x)}\, h(x).$$

Note: The natural parameter space (values of η for which this is indeed a density) is convex, and $E_\eta\{g(X)\}$ is an analytic function on the interior.

Example 1 (Bernoulli trials, $0 < p < 1$)

$$f_p(x) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}$$

$$= (1-p)^n\, e^{(\sum_1^n x_i)\log\frac{p}{1-p}} \cdot 1$$

$$= C^*(\eta)\, e^{T(x)\,\eta} \cdot 1$$

where $\eta \equiv \eta(p) \equiv \log\frac{p}{1-p} \equiv$ (log odds). (Note: $p = \frac{1}{2}$ iff $\eta = 0$.)

[Comment: Binomial(n,p) has $h(x) = \binom{n}{x}$.]

Comment. $\underset{\sim}{X} \equiv (X_1, \ldots, X_n)'$ are iid Bernoulli(p); so $T \equiv T(\underset{\sim}{X}) \equiv$ Binomial(n,p).

Then $f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = \dfrac{p^{\sum_i^n x_i}(1-p)^{n-\sum_i^n x_i}}{\binom{n}{t} p^t (1-p)^{n-t}}\, 1_{[\sum_i^n x_i = t]} = \dfrac{1}{\binom{n}{t}}\, 1_{[\sum_i^n x_i = t]}$

so that this conditional distribution is independent (is not a function of) θ. That is, T is sufficient for θ.

The dist of the suff stat is also an exponential family

Example 2. (The $N(\mu, \sigma^2)$ family) Now

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}}\right] e^{-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}} \cdot 1$$

$$= c^*(x)\, e^{\eta_1 x^2 + \eta_2 x} \cdot 1 = c^*(x)\, e^{\eta_1 T_1(x) + \eta_2 T_2(x)}.$$

Now let $X_1, \ldots X_n$ be iid $N(\mu, \sigma^2)$. Then

$$f_{\mu,\sigma^2}(\underset{\sim}{x}) = \ldots = \left[\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{n\mu^2}{2\sigma^2}}\right] e^{-\frac{\sum_1^n x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum_1^n x_i},$$

in the same spirit as before. [Note: $\mu = 0$ iff $\eta_2 = 0$ (too bad not $\frac{\mu}{\sigma}$).

Stat 512 Review (2002) — Shorack

Basic Probability

Elementary facts about probability, with partitions and Bayes rule; inclusion/exclusion
Equally likely model; and combinatorics $n!$, P_k^n, C_k^n, $\binom{n+k-1}{k-1}$, $\binom{n}{n_1,\dots,n_k}$.

Elementary expectation facts

Mean of a sum is sum of means, always.
Variance of a sum is sum of variances, if ind.
Expectation of a product is product of expectations, if ind.
So, X and Y ind implies any $g(X)$ and $h(Y)$ are independent.
$\mu \equiv EX$ = mean, $\tau \equiv E|X-\mu|$ = mean deviation, $\sigma \equiv \sqrt{E(X-\mu)^2}$ = stdev all have meaning.
One-the-hard-way calculations of many different moments.

Sampling from a Red/White urn

With replacement: $T_n \equiv X_1 + \cdots + X_n \cong Bi(n,p) \cong (np, npq)$
Without replacement: $T_n \equiv X_1 + \cdots + X_n \cong$ Hypergeometric$(R,W;n) \cong (np, npq[1-\frac{n-1}{N-1}])$; $N \equiv R+W$, $p \equiv R/N$.

Not yet

Sampling from a General Urn containing $a_1, \dots, a_N$ without replacement
One draw X has mean $\bar{a} \equiv \frac{1}{N}\sum_1^N a_i$ and variance $\sigma_a^2 \equiv \frac{1}{N}\sum_1^N (a_i - \bar{a})^2$.
$T_n \equiv X_1 + \cdots + X_n \cong (n\bar{a}, n\sigma_a^2[1 - \frac{n-1}{N-1}])$, and $Cov[X_1, X_2] = -\frac{1}{N-1}\sigma_a^2$.

Bernoulli(p) Process $X_1, X_2, \dots$ of iid Bernoulli(p) rvs

$T_n \equiv$ (total) $\equiv X_1 + \cdots + X_n \cong Bi(n,p)$
$Y_1, Y_2, \dots$ = (interarrival times) are iid GeoT$(p) \cong (\frac{1}{p}, \frac{q}{p^2})$. (Lack of memory) $(P(Y \ge h) = q^h)$
$W_r \equiv$ (waiting times) $\cong$ NegBiT(r,p).
$(T_m \mid T_n = R) \cong$ Hypergeometric$(R, n-R; m)$, regardless of the value of p.
Key Fact: $[W_r > n] = [T_n < r]$.
PLT: $T_n \cong Bi(n, p_n) \longrightarrow_d T \cong$ Poisson(λ) if $np_n \to \lambda$. [Note: $np_n \to \lambda$ = mean, $np_nq_n \to \lambda$ = variance]
Poisson is important. (It arose historically when n large, p small, both unknown.)

Poisson (intensity $= \nu$) Process $\{N_t : t \ge 0\}$

$N_t \cong$ Poisson$(\lambda = \nu t) \equiv (\lambda, \lambda)$.
$Y_1, Y_2, \dots$ = (interarrival times) are iid Exponential (rate $= \nu$) = Exponential (mean $= \theta = \frac{1}{\nu}$)
$W_r \equiv Y_1 + \cdots + Y_r$ = (waiting times) $\cong$ Gamma$(r,$ rate $= \nu)$. Lack of memory, $P(Y > t) = e^{-\nu t}$
$(N_s \mid N_t = n) \cong$ Binomial$(n, p = \frac{s}{t})$.
$(W_1, \dots, W_r \mid W_{r+1} = t) \cong$ (order stats of r iid Uniform$(0,t)$ rvs).
$(W_1, \dots, W_r \mid N_t = r) \cong$ (ditto).
So, $(\frac{W_1}{W_{r+1}}, \dots, \frac{W_r}{W_{r+1}}) \cong (U_{n:1}, \dots, U_{n:n})$; so its density is $n!$ on $0 < u_1 < \cdots < u_n < 1$.
Geiger counter, multiple light bulbs, and male vs female customers at Nordstroms.
Key Fact: $[W_r > t] = [N_t < r]$.
Poisson(λ) is approx Normal(λ, λ) for $\lambda \ge 5$
Lack of memory of $\{N_t : t \ge 0\}$; and of the Y_i's, with $P(Y > t) = e^{-\nu t}$ for $t \ge 0$.

Gamma Function $\Gamma(\cdot)$, and Beta(m,n)

Gamma$(r,$ rate $= \nu)$ and Beta(m,n) densities. ($E(\frac{1}{X})$ for these.)
Means are variance. Beta moments nearly equal certain Binomial moments
$U_{n:i} \cong$ Beta$(i, n-i+1)$ by "intuitive proof." Key Fact: $[U_{n:i} > t] = [\sum_1^n 1_{[0,t]}(U_i) < i]$

Verify $f_{N(0,1)}(\cdot)$ integrates to 1, and that $\mu = 0$ and $\sigma^2 = 1$.

Other "intuitive proof"

Convolution formula $f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\,dx = \int_{-\infty}^{\infty} f(z-y, y)\,dy$.
Marginal Beta$(i, n-i+1)$ dist of $U_{n:i}$ and joint Dirichlet dist of $U_{n:i}$ and $U_{n:j}$. not yet

Mgf's $M_X(t) \equiv Ee^{tX}$ in a neighborhood; Chf's $\phi_X(t) \equiv Ee^{itX}$ for all real t.

(1) $M_{aX+b}(t) = e^{tb} M_X(at)$ and/or $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$

(2) $M_{X+Y} = M_X M_Y$ if ind and/or $\phi_{X+Y} = \phi_X \phi_Y$ if ind

(3) Uniqueness Theorem

(4) (Cramér-Lévy continuity theorem) If $\begin{cases} M_n(t) \to M(t) \text{ in a nbhd of } 0 \\ \phi_n(t) \to \phi(t) \end{cases}$, then $F_n \to_L F$.

$M_Z(t) = e^{t^2/2}$, while $\phi_Z(t) = e^{-t^2/2}$.

Pf of CLT, based on $M_{(X_i-\mu)/\sigma\sqrt{n}}(t) = 1 + \frac{t}{\sigma\sqrt{n}} \cdot 0 + \frac{t^2\sigma^2}{n} + \frac{t}{6}\frac{t^3}{n^{3/2}}EX^3 + \cdots = 1 + \frac{a_n}{n}$

... where $a_n \to a \equiv t^2/2$.

Other mgf/chf proofs: Sums of $\begin{cases} \text{Normals} \\ \text{Cauchys} \\ \text{Gammas} \end{cases}$ or of $\begin{cases} \text{Poissons} \\ \text{Binomials} \\ \text{Neg Bi's} \end{cases}$ belong to this same family.

Limit theorems we've proved by mgf/chf:

CLT and PLT (and the merely stated FSCLT)

$pY' \to_L \text{Exp}(1)$ as $p \to 0$, and $pW_r' \to_L \text{Gamma}(r,1)$ as $p \to 0$.

$W_r' \to_L \text{Poisson}(\lambda)$ as $r \to \infty$ provided $r(1-p_r) \to \lambda$.

$X - \mu_X \Rightarrow$ lever arm, distance from μ_X

$Y - \mu_X \Rightarrow$ lever arm, distance from μ_Y

EXPECTATION

Definition 1: Random outcome X, Y yields payoff $g(X,Y)$, with average value

$$Eg(X,Y) \equiv \begin{cases} \sum\sum_{\text{all } x,y} g(x,y)\, p_{X,Y}(x,y) & \text{(discrete rv's)} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\cdot f_{X,Y}(x,y)\,dx\,dy & \text{(continuous rv's)} \end{cases}$$

Example 1.

$\mu_X \equiv EX$ and $\sigma_X^2 \equiv E(X-\mu_X)^2$; μ_X and σ_X measure center and spread of X.
$\mu_Y \equiv EY$ and $\sigma_Y^2 \equiv E(Y-\mu_Y)^2$; μ_Y and σ_Y measure center and spread of Y.

$\sigma_{X,Y} \equiv \text{Cov}[X,Y] \equiv E\{(X-\mu_X)(Y-\mu_Y)\}$ and $\rho_{X,Y} \equiv \text{Corr}[X,Y] \equiv \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ measure the relationship between X and Y.

Theorem 1.

$E\{a\} = a$
$E\{aX+b\} = aE\{X\}+b$ or $\mu_{aX+b} = a\mu_X + b$
$E\{aX+bY+c\} = aE\{X\}+bE\{Y\}+c$ always or $\mu_{aX+bY+c} = a\mu_X + b\mu_Y + c$ always
$E\{X+Y\} = E\{X\}+E\{Y\}$ always or $\mu_{X+Y} = \mu_X + \mu_Y$ always

$\Leftrightarrow$
$\text{Var}[aX+b] = a^2 \text{Var}[X]$ or $\sigma_{aX+b} = |a|\sigma_X$
$\text{Var}[X] = 0$ if and only if X only assumes one value (μ_X must equal that value)

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ if X and Y are independent; or $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$ if X and Y are independent.
$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}[X,Y]$ always ; or $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$ always.
$\text{Var}[aX+bY+c] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\,\text{Cov}[X,Y]$; or $\sigma_{aX+bY+c}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\sigma_{X,Y}$.

$\Leftrightarrow$
Computing formulas:
$\text{Var}[X] = EX^2 - E^2\{X\}$; or $\sigma_X^2 = EX^2 - \mu_X^2$ (or $EX^2 = \sigma_X^2 + \mu_X^2$).
$\text{Cov}[X,Y] = E(XY) - EX\,EY$; or $\sigma_{X,Y} = E(XY) - \mu_X\mu_Y$.
$\text{Var}[X] = \text{Cov}[X,X]$.
$\text{Cov}[X,Y] = 0$ whenever X and Y are independent.

$\Leftrightarrow$
The key fact in all this: $\int g(x) f_X(x)\,dx \cdot \int h(y) f_Y(y)\,dy$
$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\}$ for any $g(\cdot)$ and $h(\cdot)$ whenever X and Y are independent.
$\Leftarrow$ Thus $E\{(X-\mu_X)(Y-\mu_Y)\} = E(X-\mu_X)E(Y-\mu_Y) = 0\cdot 0 = 0$ when X and Y are independent.

General Form $\text{Cov}\left[\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right] = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}[X_i, Y_j]$ since $\left[\sum_{i=1}^m a_i(X_i-\mu_{X_i})\right]\left[\sum_{j=1}^n b_j(Y_j-\mu_{Y_j})\right] = \sum_{i=1}^m\sum_{j=1}^n a_i b_j (X_i-\mu_{X_i})(Y_j-\mu_{Y_j})$

$b_i = a_i$

$\text{Var}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 \text{Var}[X_i] + \sum_{i\neq j}\sum a_i a_j \text{Cov}[X_i, X_j]$.

su
$a = 1$

$\text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i] + \sum_{i\neq j}\sum \text{Cov}[X_i, X_j]$

$= n\,\text{Var}[X_1] + n(n-1)\,\text{Cov}[X_1, X_2]$ if all variances and covariances are equal.

- In general, if Cov = 0 does not mean independence

Example 2. (Urn mean and variance) An urn contains N balls numbered $a_1, \ldots, a_N$. Let X denote the a-value of one ball drawn at random from the urn.

$\mu_X = \bar{a} \equiv \frac{1}{N}\sum_{i=1}^N a_i$ and $\sigma_X^2 = \sigma_a^2 \equiv \frac{1}{N}\sum_{i=1}^N (a_i - \bar{a})^2 = \frac{1}{N}\sum_{i=1}^N a_i^2 - \bar{a}^2$

Example 3. (Sampling with replacement from an urn) Let $T_n \equiv X_1 + \cdots + X_n$. Then

$\mathrm{Var}[T_n] = n\sigma_a^2$ or $\therefore \mathrm{Var}[\bar{X}_n] = \frac{1}{n}\sigma_a^2$ . . $\mathrm{Cov}[X_1, X_2] = 0$

Example 4. (Sampling without replacement from an urn) Let $T_n \equiv X_1 + \cdots + X_n$. Then

$\mathrm{Var}[T_n] = n\sigma_a^2\left(1 - \frac{n-1}{N-1}\right)$ or $\mathrm{Var}[\bar{X}_n] = \frac{1}{n}\sigma_a^2\left(1 - \frac{n-1}{N-1}\right)$. $\therefore \mathrm{Cov}[X_1, X_2] = -\frac{\sigma_a^2}{N-1}$.

The correction factor $\left(1 - \frac{n-1}{N-1}\right)$ is due to the covariance no longer being 0. It has the same value no matter what numbers $a_1, \ldots, a_N$ are in the urn.

Proof. $0 = \mathrm{Var}[X_1 + \cdots + X_N] = N\sigma_a^2 + N(N-1)\mathrm{Cov}[X_1, X_2]$, so $\mathrm{Cov}[X_1, X_2] = -\frac{\sigma_a^2}{N-1}$.

Then $\mathrm{Var}[X_1 + \cdots + X_n] = n\sigma_a^2 + n(n-1)\mathrm{Cov}[X_1, X_2] = n\sigma_a^2 + n(n-1)\left(-\frac{\sigma_a^2}{N-1}\right) = n\sigma_a^2\left[1 - \frac{n-1}{N-1}\right]$. □

The Central Limit Theorem (CLT)

Independent Repetitions (Sampling With Replacement is a special case of this).

The basic X-experiment has mean μ and standard deviation σ.
Let $T_n \equiv X_1 + \cdots + X_n$ for independent repetitions $X_1, \ldots, X_n$.

T_n has mean $n\mu$ and standard deviation $\sqrt{n}\,\sigma$.
$\bar{X}_n$ has mean μ and standard deviation $\sigma/\sqrt{n}$.

$Z_n \equiv \frac{T_n - n\mu}{\sqrt{n}\,\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ has mean 0 and standard deviation 1.

CLT: The distribution of Z_n converges to the $N(0,1)$ distribution.

Sampling Without Replacement from an $a_1, \ldots, a_N$ urn with $\mu = \bar{a} \equiv \frac{1}{N}\sum_{i=1}^{N} a_i$ and $\sigma^2 = \sigma_a^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a})^2$.

One draw X from the urn has mean $\mu = \bar{a}$ and standard deviation σ_a.
Let $T_n \equiv X_1 + \cdots + X_n$ for the dependent repetitions $X_1, \ldots, X_n$.

T_n has mean $n\bar{a}$ and standard deviation $\sqrt{n}\,\sigma_a\sqrt{1 - \frac{n-1}{N-1}}$.
$\bar{X}_n$ has mean $\bar{a}$ and standard deviation $\frac{1}{\sqrt{n}}\sigma_a\sqrt{1 - \frac{n-1}{N-1}}$.

$Z_n \equiv \frac{T_n - n\bar{a}}{\sqrt{n}\,\sigma_a\sqrt{1 - \frac{n-1}{N-1}}} = \frac{\bar{X}_n - \bar{a}}{\frac{1}{\sqrt{n}}\sigma_a\sqrt{1 - \frac{n-1}{N-1}}}$ has mean 0 and standard deviation 1.

FSCLT: The distribution of Z_n converges to the $N(0,1)$ distribution provided n and $N-n$ both $\to \infty$ and $\max_{1 \le i \le N} \frac{|a_i - \bar{a}|}{\sqrt{N}\,\sigma_a} \to 0$

Shorack STAT 506 April 21, 2004

First Midterm (closed book and closed notes).

#1. (25) The complete roster of the View Ridge soccer club has 4 attackers, 6 midfielders, 6 defenders, and 2 goalies. How many different teams of 3 attackers, 3 midfielders, 4 defenders, and 1 goalie could be formed in order to play another soccer club.

#2. (25) **4.** Two customers per minute arrive at the checkout of Matthew's Metro Market.
(a) What is the probability that at least 2 customers arrive at the checkout area in the next two minutes?
(b) What is the probability that at least 100 customers arrive at the checkout area in the next hour?
(c) Given that exactly 100 customers arrive in the next hour, what is the probability that at least 20 of them arrive in the second 15 minutes?

#3. (25) Suppose X and Y have the joint density

$$f(x,y) = 2e^{-2x}\frac{1}{x} \qquad \text{on } 0 \le y \le x < \infty.$$

(a) Determine the marginal density $f_X(\cdot)$ of X. Name this familiar density.
(b) Determine the conditional density $f_{Y|X=x}(\cdot)$ of Y given that $X = x$. State the name of this familiar density. Because it is familiar, you can immediately write out formulas for the conditional mean $m(x) \equiv \mathrm{E}\{Y \mid X = x\}$ and the conditional variance $v^2(x) \equiv \mathrm{Var}[Y \mid X = x]$. Do so. (Calculation now is not necessary.)
(c) Evaluate μ_Y.

#4. (25) (a) Compute the value $M(t)$ of the moment generating function for each possible t when Y is a Poisson(λ) rv. (This is a problem you have done before.)
(b) Use this mgf to compute $\mu = \mathrm{E}(Y)$. (No other method is allowed.)

THE EXAM IS OVER. THIS IS HOMEWORK–to be discussed Monday.

#1.(cont.)
(b) How many choices are possible for an "A team" of 2 attackers, 3 midfielders, 3 defenders, and 1 goalie that will play a practice game against a "B team" comprised of the remaining players.

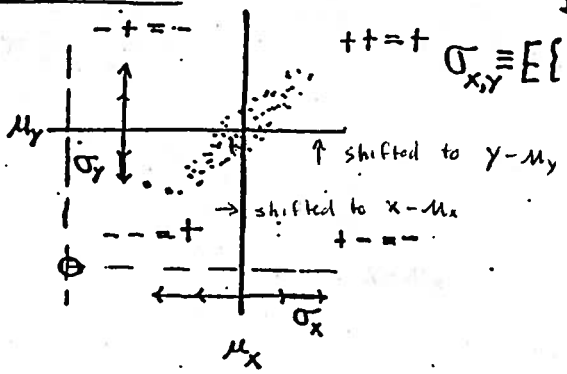
#2. (cont.)
(d) Break the upcoming hour into six disjoint ten minute intervals. We say that a "customer overload" occurs during this hour if at least 20 customers arrive during any one of these 6 ten minute intervals. What is the probability of a customer overload during the upcoming hour?
(e) What is the probability that the second customer overload comes in the sixth ten minute interval?

#3. (cont.)
(d) Determine μ_X, σ_X^2, μ_Y, σ_Y^2, and $\rho_{X,Y}$.
(e) Determine the median $\ddot{m}(x)$ of the conditional distribution of Y given $X = x$.

#4.(cont.)
(c) Use $M(\cdot)$ to compute σ^2.
(d) Use $\log M(\cdot)$ to compute σ^2.
(e) Evaluate σ^2 via a "one the hard way" proof. (You already know the value of μ.)

Correlation & Covariance

Comment on the meaning of covariance.



$$\sigma_{X,Y} \equiv E\{(X-\mu_X)\cdot(Y-\mu_Y)\} = E\left\{\begin{matrix}+ & +\\ - & -\end{matrix} : \begin{matrix}+ & -\\ - & +\end{matrix}\right\} = E\{+\}$$

= (tends to be positive in this type of picture)

$$\mathrm{Cov}\left[\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right] = E\left(\frac{X-\mu_X}{\sigma_X}\,\frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y} \equiv \rho_{X,Y} \equiv \rho$$

is just the covariance of the standardized rv's $\frac{X-\mu_X}{\sigma_X}$ and $\frac{Y-\mu_Y}{\sigma_Y}$.

Thus ρ does not depend on the units in which X and Y are measured.
Thus correlation is a much more useful measure of relationship, than is $\sigma_{X,Y}$

Theorem 2. (The correlation inequality).

$-1 \le \rho \le 1$ always.

$\rho = 0$ when X and Y are independent.

$\rho = \pm 1$ implies $Y-\mu_Y = \pm\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$.

Proof

$$0 \le \mathrm{Var}\left[\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y}\right] = \frac{\mathrm{Var}[X]}{\sigma_X^2} + \frac{\mathrm{Var}[Y]}{\sigma_Y^2} \pm 2\frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y}$$

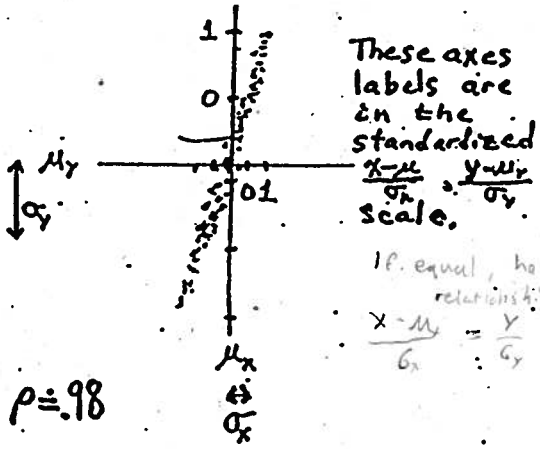
$$= 1+1\pm 2\rho = 2(1\pm\rho).$$

$0 \le 1+\rho$ implies $-1 \le \rho$, while

$0 \le 1-\rho$ implies $\rho \le 1$.

$0 = \mathrm{Var}$ if and only if $\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y} = 0$.

Thus, ρ measures linear dependence.

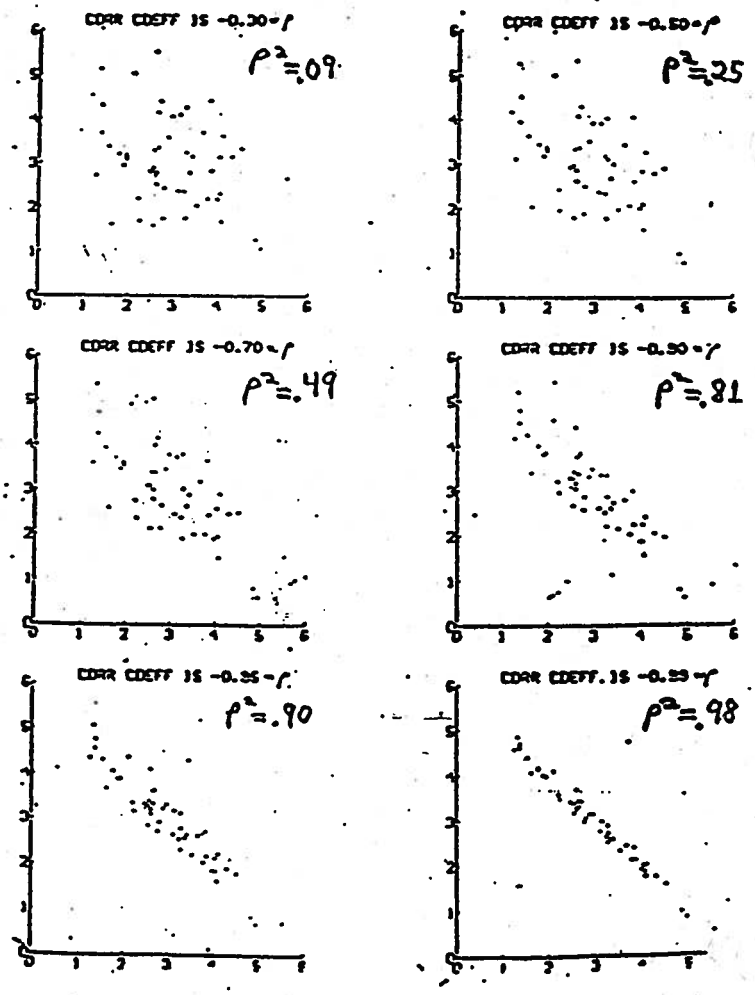


These axes labels are in the standardized $\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}$ scale.

If equal, have linear relationship.

$$\frac{X-\mu_X}{\sigma_X} = \frac{Y}{\sigma_Y} - \frac{\mu_Y}{\sigma_Y}$$

$\rho \doteq .98$


CORR COEFF IS −0.30 = ρ
ρ² = .09
CORR COEFF IS −0.50 = ρ
ρ² = .25
CORR COEFF IS −0.70 = ρ
ρ² = .49
CORR COEFF IS −0.90 = ρ
ρ² = .81
CORR COEFF IS −0.95 = ρ
ρ² = .90
CORR COEFF. IS −0.99 = ρ
ρ² = .98

ρ only show strength of linear relationship

If strong, nonlinear relationship exists,

"PLANNED" Syllabus for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Ch1: Probability, conditional probability, independence, and rvs
Mar 31 L2...Sec2.1–2.3$^-$: E(·), mean, variance, and (especially) transformations (no mgf)
Apr 01 Q....HW and some catchup lecturing
L3...Sampling with and without replacement
.....
Apr 05 L4...Sec3.1–3.2(Handout): Bernoulli trials; Geometric, NegBinomialT (and Hypergeometric and Discrete Uniform)
L5...Sec3.3(Handout): Poisson process, Poisson, Exponential, Gamma (and Beta and Uniform)
Q ....HW, and finish Poisson process
L6...Sec2.3: Moment generating function (mgf)
.....
Apr 12 L7...More mgf, and CLT pics
L8...Ch4: Joint, marginal, and conditional distributions
Q....Sec4.1: Conditionals and convolution
L9...Sec4.2: Conditional mean and variance
.....
Apr 19 L10..Sec4.2 Independence and Expectation; Beta densities
L11..Sec3.4 Exponential families; list of variance and covariance formulas
Q....MIDTERM I
L12..Discuss Midtem; the intuitive meaning of covariance; more on the var/cov list
.....
Apr 26 L13..Sec4.5: Correlation inequaliaty; the empirical df; finish the var/cov list
L14..Sec5.1–5.2: Properties of $\bar{X}$ and S^2
Q....
Apr 30 L15..Sec4.5: Bivariate Normal Distribution
.....
May 03 L16..Sec5.3: Derivation of Student-T and Fisher/Snedecor/varianceRatio-F
L17..Handout: Moments via matrices
Q....
L18..Sec7.1–7.2:Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)
.....
May 10 L19..MLE: Three illustrative MLE examples
L20..Sec6.1–6.2: Sufficiency
Q....Work HW on MLE
L21..Completenes, Exponential families, ancillarity
.....
May 17 L22..Bayes estimators
L23..Bayes examples
Q....MIDTERM II
L24..Discuss Midterm, and begin anew
.....
May 24 L25..Crameŕ–Rao inequality
L26..Finish Crameŕ–Rao
Q....
L27..Hypothesis tests
.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28..Likelihood ratio tests
Q....Work examples of LR-tests
June04 L29..Location and scale models
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

Planned Homework for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29	L1...Chapter1: #[2,5,11,13,14,25,33,38; do not be formal; should be easy];#17,18,21,23,24
Mar 31	L2...Chapter1: #44,53; Chapter2: #4,8c,17,19,24
Apr 01	Q....
	L3...Chapter2: #20; Chapter3: #4,6,12,23,24b; Handout: #1
.....	
Apr 05	L4...Handout: #2–5
	L5...Handout: #6–7
	Q ...
	L6...
.....	
Apr 12	L7...Chapter2: #9,16, (read 11); Chapter3: #22cd
	L8...Chapter2: #33abc; Chapter3: #19,22ab
	Q....
	L9...Chapter2: # 26ab, 28ab, 30ad; Chapter3: #24c
.....	
Apr 19	L10..Chapter4 #13,15; HW #1abc in Conditional Epectation handout about $v^2(y)$
	L11..Unoficial: Work some off the HW sheet containing nine
	Q....MIDTERM
	L12..HW on the back of the midterm
.....	
Apr 26	L13..Chapter4 #39, 43, 58
	L14..Chapter4 #23(long, re transformations–start early), 44
	Q....
Apr 30	L15..Chapter4 # 27, 30, 45
.....	
May 03	L16..Chapter4 #19, 20(discuss #21 as well as a certain in class example in this context), 54
	L17..Chapter4 #42, 52
	Q....
	L18..
.....	
May 10	L19
	L20
	Q....
	L21..
.....	
May 17	L22..
	L23..
	Q....
	L24..
.....	
May 24	L25..
	L26..
	Q....
	L27..
.....	
May 31	MEMORIAL DAY HOLIDAY
June02	L28..
	Q
June04	L29..
.....	
June09	FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....	

Problems

1. The National Center for Educational Statistics recently chose a random sample of 2000 newly graduated college students and queried each one about the time it took to complete his or her degree. If the sample mean was 5.2 years with a sample standard deviation of 1.2 years, construct
   (a) A 95 percent confidence interval estimate of the mean completion time of all newly graduated students
   (b) A 99 percent confidence interval estimate

2. The manager of a shipping department of a mail-order operation located in New York has been receiving complaints about the length of time it takes for customers in California to receive their orders. To learn more about this potential problem, the manager chose a random sample of 12 orders and then checked to see how many days it took to receive each of these orders. The resulting data were

   15, 20, 10, 11, 7, 12, 9, 14, 12, 8, 13, 16

   (a) Find a 90 percent confidence interval estimate for the mean time it takes for California customers to receive their orders.
   (b) Find a 95 percent confidence interval estimate.

3. A survey was instituted to estimate μ, the mean salary of middle-level bank executives. A random sample of 15 executives yielded the following yearly salaries (in units of $1000):

   88, 121, 75, 39, 52, 102, 95, 78, 69, 82, 80, 84, 72, 115, 106

   Find a
   (a) 90 percent
   (b) 95 percent
   (c) 99 percent
   confidence interval estimate of μ.

4. The numbers of riders on an intercity bus on 12 randomly chosen days are

   47, 66, 55, 53, 49, 65, 48, 44, 50, 61, 60, 55

   (a) Estimate the mean number of daily riders.
   (b) Estimate the standard deviation of the daily number of riders.
   (c) Give a 95 percent confidence interval estimate for the mean number of daily riders.

5. Use the data of Prob. 1 of Sec. 8.2 to obtain a
   (a) 95 percent
   (b) 99 percent
   confidence interval estimate of the average weight of all participants of the 1990 Boston Marathon.

6. A random sample of 30 General Electric transistors resulted in an average lifetime of 1210 hours with a sample standard deviation of 92 hours. Compute a
   (a) 90 percent
   (b) 95 percent
   (c) 99 percent
   confidence interval estimate of the mean life of all General Electric transistors.

7. In Prob. 10 of Sec. 8.4 determine a 95 percent confidence interval estimate of the population mean burn temperature.

8. The following are the losing scores in seven randomly chosen Super Bowl football games:

   10, 16, 20, 17, 31, 19, 14

   Construct a 95 percent confidence interval estimate of the average losing score in Super Bowl games.

9. The following are the winning scores in eight randomly chosen Masters Golf Tournaments:

   285, 279, 280, 288, 279, 286, 284, 279

   Use these data to construct a 90 percent confidence interval estimate of the average winning score in the Masters.

10. All the students at a certain school are to be given a psychological task. To determine the average time it will take a student to perform this task, a random sample of 20 students was chosen and each was given the task. If it took these students an average of 12.4 minutes with a sample standard deviation of 3.3 minutes, find a 95 percent confidence interval estimate for the average time it will take all students in the school to perform this task.

11. A large company self-insures its large fleet of cars against collisions. To determine its mean repair cost per collision, it has randomly chosen a sample of 16 accidents. If the average repair cost in these accidents is $2200 with a sample standard deviation of $800, find a 90 percent confidence interval estimate of the mean cost per collision.

12. An anthropologist measured the heights (in inches) of a random sample of 64 men of a certain tribe, and she found that the sample mean was 72.4 and the sample standard deviation was 2.2. Find a
    (a) 95 percent
    (b) 99 percent
    confidence interval estimate of the average height of all men of the tribe.

13. To determine the average time span of a phone call made during midday, the telephone company has randomly selected a sample of 1200 such calls. The sample mean of these calls is 4.7 minutes, and the sample standard deviation is 2.2 minutes. Find a
    (a) 90 percent (b) 95 percent
    confidence interval estimate of the mean length of all such calls.

14. Each of 20 science students independently measured the melting point of lead. The sample mean and sample standard deviation of these measurements were (in degrees Celsius) 330.2 and 15.4, respectively. Construct a
    (a) 95 percent (b) 99 percent
    confidence interval estimate of the true melting point of lead.

15. A random sample of 300 CitiBank VISA cardholder accounts indicated a sample mean debt of $1220 with a sample standard deviation of $840. Construct a 95 percent confidence interval estimate of the average debt of all cardholders.

16. To obtain information about the number of years that Chicago police officers have been on the job, a sample of 46 officers was chosen. Their average time on the job was 14.8 years with a sample standard deviation of 8.2 years. Determine a
    (a) 90 percent
    (b) 95 percent
    (c) 99 percent
    confidence interval estimate for the average time on the job of all Chicago police officers.

The Sample Mean $\bar{X}_n$ and Sample Variance S_n^2

GRS

Let $X_1, \ldots, X_n$ be an iid sample having mean μ and variance σ^2.

However, in statistical problems, μ and σ are unknown. Use:

$\bar{X} \equiv \bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ and $T_n \equiv \sum_{i=1}^n X_i$.

$\hat{\sigma}_n^2 \equiv \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ and $\hat{\sigma}_{on}^2 \equiv \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 = \bar{Y}_n$ if $Y_i \equiv (X_i - \mu)^2$

$= $ (unobservable)

For indep random variable $E(Y_i)^2 = \mu_4 = E(X_i - \mu)^4$

Fact 1. $\bar{X}_n \cong (\mu, \frac{\sigma^2}{n})$ Mean is unbiased

Fact 2. $ES_n^2 = \sigma^2$ and $\mathrm{Var}[S_n^2] = \frac{1}{n}\{\mu_4 - \frac{n-3}{n-1}\sigma^4\} = \frac{1}{n}\{\mu_4 - \sigma^4\} + \frac{2\sigma^4}{n(n-1)}$

where $\mu_4 \equiv E(X-\mu)^4$ is the fourth central moments of X. (See Fact 6.)

Fact 3. (Identity) $\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2$

which is minimized by the choice $a = \bar{x}$.

effect of $\bar{x}$ instead of μ goes to 0 as $n \to \infty$

Fact 4. (Recall) $E(X-a)^2 = E(X-\mu)^2 + (\mu - a)^2 = \sigma^2 + (\mu - a)^2 \geq \sigma^2$

which is minimized only by $\mu = a$.

[Let the empirical df $F_n(\cdot)$ put mass $\frac{1}{n}$ at each x_i for $1 \leq i \leq n$.]

Fact 5. (Recall) $E(Y - g(X))^2 = E(Y - m(X))^2 + E(m(X) - g(X))^2 \geq E(Y - m(X))^2 = E v^2(X)$

where $m(x) \equiv E(Y|X=x)$ is the mean of $F_{Y|X=x}(\cdot)$ and $v^2(x)$ is its variance

[Equality holds iff $g(x) = m(x)$. The minimized value is $E v^2(X)$.]

Note. Apply Fact 4 to the empirical distribution (which has df F_n and mean $\bar{X}_n$) gives

$$\frac{1}{n}\sum_1^n (x_i - a)^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2 + (\bar{x} - a)^2 = \hat{\sigma}_n^2 + (\bar{x} - a)^2 \geq \hat{\sigma}_n^2 \equiv \frac{SS_{xx}}{n}.$$

Applying the correlation inequality $[\mathrm{Cov}[X,Y]]^2 \leq \sigma_x^2 \sigma_y^2$ to the empirical dist that put mass $\frac{1}{n}$ at each data point $(x_1, y_1), \ldots, (x_n, y_n)$ gives

$$\frac{1}{n}SS_{xy} \equiv \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \leq \hat{\sigma}_x^2 \hat{\sigma}_y^2 = \frac{1}{n}SS_{xx} \cdot \frac{1}{n}SS_{yy} = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2 \cdot \frac{1}{n}\sum_1^n (y_i - \bar{y})^2.$$

Thus $-1 \leq \hat{\rho} \leq 1$ for $\hat{\rho} \equiv \frac{\frac{1}{n}SS_{xy}}{\sqrt{\frac{1}{n}SS_{xx}\,\frac{1}{n}SS_{yy}}} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{\frac{1}{n-1}SS_{xy}}{\sqrt{\frac{1}{n-1}SS_{xx} \cdot \frac{1}{n-1}SS_{yy}}} \equiv \frac{S_{xy}}{S_x S_y}$.

Fact 6. $\hat{\sigma}_{on}^2$ has mean σ^2 and variance $\frac{1}{n}(\mu_4 - \sigma^4)$.

Fact 7. $\mathrm{Cov}[\bar{X}, S^2] = \frac{\mu_3}{n}$ where $\mu_3 \equiv E(X-\mu)^3$.

Exercise Prove Facts 6 and 7.

How can we compute the distribution of $\bar{X}_n$?

(#1) **Convolution formula**: $Z = X+Y$ with X and Y independent.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \qquad \text{(continuous)}$$

$$f_Z(z) = \sum_{\text{all } x} f_X(x) f_Y(z-x) \qquad \text{(discrete)}$$

(#2) **Moment generating functions** (characteristic functions):

$$M_Z(t) = M_X(t)\,M_Y(t)$$ ← used to prove theorems

(#3) **CLT**: $\dfrac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{d} Z \cong N(0,1)$

That is, $Z_n = \sqrt{n}(\bar{X}_n-\mu)/\sigma \cong (0,1)$ satisfies

$$P(\sqrt{n}(\bar{X}_n-\mu)/\sigma \le z) = P(Z_n \le z) \to P(Z \le z) \text{ for all } z;$$

and we write $Z_n \xrightarrow{d} Z$.

WLLN A much weaker result is $\bar{X}_n \xrightarrow{p} \mu$; meaning

$P(|\bar{X}_n-\mu| \ge \epsilon) \to 0$ as $n \to \infty$, for each $\epsilon > 0$.

↳ Probability that $\bar{X}_n$ is in interval $(\mu_X-\epsilon, \mu_X+\epsilon)$

Pf. By Chebyshev, $P(|\bar{X}_n-\mu| \ge \epsilon) \le \frac{1}{\epsilon^2}\text{Var}[\bar{X}_n] = \frac{\sigma_X^2}{\epsilon^2 n} \to 0$ for all $\epsilon > 0$. ∎

Says a large sample converges to this value; always true but poor approx.

Cor $\hat{\sigma}_{on}^2 \equiv \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^n Y_i = \bar{Y}_n \xrightarrow{p} \mu_Y = \sigma_X^2$.

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2 - (\bar{X}_n-\mu)^2 = \hat{\sigma}_{on}^2 - (\bar{X}_n-\mu)^2 \to \sigma_X^2 - 0^2 = \sigma_X^2.$$

$$S_n^2 = \frac{n-1}{n}\hat{\sigma}_n^2 \xrightarrow{p} 1 \cdot \sigma_X^2 = \sigma_X^2.$$

Thus, $\hat{\sigma}_n^2$ and S_n^2 are both consistent estimators of σ^2.

Just as $\bar{X}_n$ is a consistent estimator of μ.

(Note: we showed in the corollary that $\hat{\sigma}_n^2 = \hat{\sigma}_{on}^2 - (\bar{X}_n-\mu)^2$ and that $S_n^2 = \hat{\sigma}_n^2 - \frac{1}{n}\hat{\sigma}_n^2$.

CLT. But more is true.

$$\sqrt{n}(\hat{\sigma}_{on}^2 - \sigma^2) = \sqrt{n}(\bar{Y}-\mu_Y) \xrightarrow{d} N(0,\sigma_Y^2) = N(0,\mu_4-\sigma^4) \text{ where } \mu_4 \equiv E(X-\mu)^4.$$

So $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \sqrt{n}(\hat{\sigma}_{on}^2 - \sigma^2) - \frac{1}{\sqrt{n}}[\sqrt{n}(\bar{X}_n-\mu)]^2 \xrightarrow[\text{Slutsky thm}]{d} N(0,\mu_4-\sigma_X^4)$

and $\sqrt{n}(S_n^2 - \sigma^2) = \sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) - \sqrt{n}\left(\frac{1}{n}\hat{\sigma}_n^2\right) \xrightarrow[\text{Slutsky thm}]{d} N(0,\mu_4-\sigma_X^4)$

OR $\dfrac{\sqrt{n}(S_n^2-\sigma^2)}{\sqrt{2}\,\sigma^2} \to N\left(0, \dfrac{E(X-\mu_X)^4-\sigma^4}{2\sigma^4}\right) = N\left(0, 1+\dfrac{\gamma_2}{2}\right)$ where $\gamma_2 \equiv \dfrac{E(X-\mu)^4-3\sigma^4}{\sigma^4}$

$= \{$ dimensionless $=0$ under normality

Random Sampling From an Urn (i.e. Any experiment)

Basic Experiment X is one draw from the urn.
Let μ denote the population mean.
Let σ denote the population stdev.

Let $X_1, \dots, X_n$ denote repetitions of the expmt; it's the sample.

$T_n \equiv X_1 + \cdots + X_n$ is the sample total

$\bar{X}_n \equiv \frac{X_1 + \cdots + X_n}{n} = \frac{T_n}{n}$ is the sample mean.

$S_n \equiv \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$ is the sample stdev

We accept that:
$\bar{X}_n$ estimates μ
S_n estimates σ
And, they work better and better as n grows,
But, how good are they. We want to quantify this.

Fact: $\mu_{\bar{X}_n} = \mu$ always, $\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}$ when independent,

$\mu_{S_n^2} = \sigma^2$ when independent average value of S

Sampling with Replacement ($X_1, \dots, X_n$ are independent)

Then T_n has mean $n\mu$ and stdev $\sqrt{n}\,\sigma$,

and $\bar{X}_n = \frac{T_n}{n}$ has mean μ and std $\sigma/\sqrt{n}$.

CLT of Prob: $\frac{T_n - n\mu}{\sqrt{n}\,\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \equiv Z_n$ is approx $Z \sim N(0,1)$.

{How normal is it? Remember our pictures!
And, remember our principles! No giants among peons.
We want a lot of 'mixing' going on.} tells us where $\bar{X}$ lives

Moreover, S_n is close to σ. Thus:

CLT of Stat: $\frac{T_n - n\mu}{\sqrt{n}\,S_n} = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$ is approx Z. tells us where μ lives

But (even better) it is approx Student's t_{n-1}!

CLT of Prob: μ and σ are known. It tells us about $\bar{X}_n$.

CLT of Stat: $\bar{X}_n$ and S_n can be observed. It tells us about μ.

Sampling Without Replacement ($X_1, \ldots, X_n$ are dependent; but, we can calculate the sum)

Then T_n has mean $n\mu$ and stdev $\sqrt{n}\,\sigma\sqrt{1-\frac{n-1}{N-1}}$.

and $\bar{X}_n$ has mean μ and stdev $\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n-1}{N-1}}$.

Both the CLT of Prob and the CLT of Stat still hold; i.e.:

$\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n-1}{N-1}}}$ is approx Z. $\frac{\bar{X}_n-\mu}{\frac{S_n}{\sqrt{n}}\sqrt{1-\frac{n-1}{N-1}}}$ is approx Students t_{n-1}.

Example 1 (Success/Failure Expmts)

● R reds, $p=\frac{R}{N}$
○ W whites, $q=\frac{W}{N}$
$N=R+W$ total

(a) **With replacement**

$\bar{X}_n$ has mean $\mu=p$ and stdev $\frac{\sigma}{\sqrt{n}}=\sqrt{\frac{pq}{n}}$.

(b) **Without replacement**

$\bar{X}_n$ has mean $\mu=p$ and stdev $\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n-1}{N-1}}=\sqrt{\frac{pq}{n}}\sqrt{1-\frac{n-1}{N-1}}$

Example 2. (Point Count in Bridge) You get 13 cards at random from the deck of:

4 Aces at 4 points each
4 Kings at 3 points each
4 Queens at 2 points each
4 Jacks at 1 point each
36 Others at 0 points each

X represents one draw from this urn.

$\mu=\left(4\cdot\frac{4}{52}\right)+\left(3\times\frac{4}{52}\right)+\left(2\times\frac{4}{52}\right)+\left(1\times\frac{4}{52}\right)+\left(0\times\frac{36}{52}\right)$

$EX^2=\left(4^2\times\frac{4}{52}\right)+\left(3^2\times\frac{4}{52}\right)+\left(2^2\times\frac{4}{52}\right)+\left(1^2\times\frac{4}{52}\right)+\left(0^2\times\frac{36}{52}\right)$

So $\mu=\frac{40}{52}=.7692$, $EX^2=\frac{120}{52}=2.308$, $\sigma=\sqrt{2.308-(.7692)^2}=1.310$.

Thus T_{13} has mean $n\mu=13\times\frac{40}{52}=10$ and

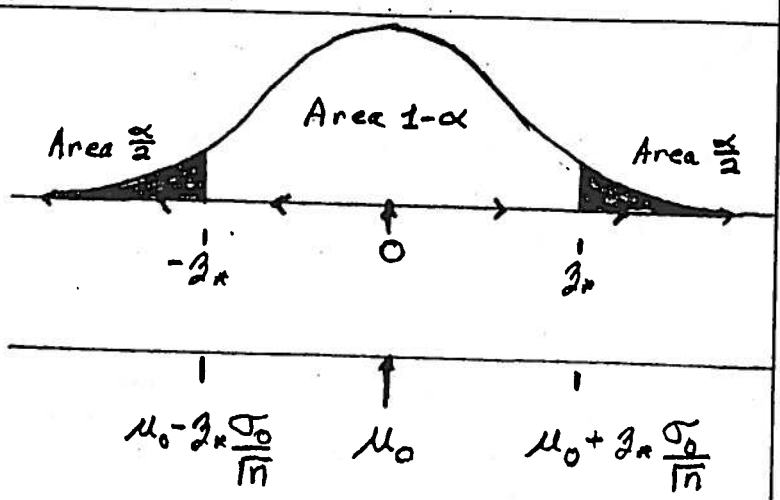
and stdev $=\sqrt{n}\,\sigma\sqrt{1-\frac{n-1}{N-1}}=\sqrt{13}\times1.310\times\sqrt{1-\frac{12}{51}}=4.131$.

Thus

$$P(T_{13}\geq 15)=P\left(\frac{T_{13}-\mu_T}{\sigma_T}\geq\frac{14.5-10}{4.131}\right)\doteq P(Z\geq 1.089)=.138.$$

$$P(T_{13}\geq 17)=P\left(\frac{T_{13}-\mu_T}{\sigma_T}\geq\frac{16.5-10}{4.131}\right)\doteq P(Z\geq 1.573)=.0579$$

$1-\alpha$	$\frac{\alpha}{2}$	z_*
.990	.005	2.576
.950	.025	1.960
.900	.050	1.645



CLT of Prob (Suppose the mean μ_0 and stdev σ_0 of X are known)

Then $\dfrac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}$ has (approximately) the Z distribution

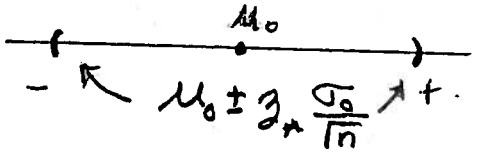
or $N(0,1)$

So $1-\alpha \doteq P\left(-z_* \le \dfrac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} \le z_*\right)$

$= P\left(-z_* \dfrac{\sigma_0}{\sqrt{n}} \le \bar{X}-\mu_0 \le z_* \dfrac{\sigma_0}{\sqrt{n}}\right)$

$= P\left(\mu_0 - z_* \dfrac{\sigma_0}{\sqrt{n}} \le \bar{X} \le \mu_0 + z_* \dfrac{\sigma_0}{\sqrt{n}}\right)$

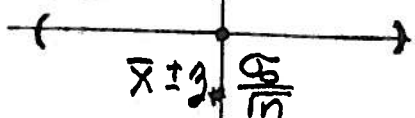
probability That is $\bar{X}$ falls in the fixed known interval $\mu_0 \pm z_* \dfrac{\sigma_0}{\sqrt{n}}$ with probability approximately equal to $1-\alpha$.



(prediction interval)

That is, the population values μ_0 and σ_0 give info about $\bar{X}$.

almost statistics **OR** See (*) $1-\alpha \doteq P\left(\bar{X} - z_* \dfrac{\sigma_0}{\sqrt{n}} \le \mu \le \bar{X} + z_* \dfrac{\sigma_0}{\sqrt{n}}\right)$

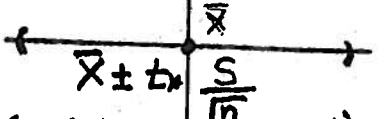


CLT of Stat (Suppose the mean μ and stdev σ of X are unknown)

Then $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ has (approximately) the t distribution.

student's $n-1$

statistics That is μ is covered by the random by with probability approximately equal to $1-\alpha$.



(Confidence interval)

- Central Limit Theorem Of Probability

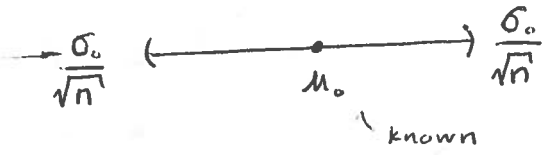
$$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma_0} \longrightarrow Z \cong N(0,1)$$

μ_0, σ_0 — known parameters

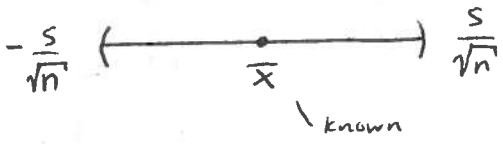
- Central Limit Theorem Of Statistics

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \longrightarrow T_{n-1} \to Z$$

- Probability: Estimating $\bar{X}$ ← unknown



- Statistics: Estimating μ ← unknown



- Theoretical confidence interval
  - $\bar{X}$, S are random variables (looking into future, experiment not performed yet)
  - confidence interval contains unknown parameter with high probability
  - After one experiment done

    $\bar{X} \to \bar{x}$ $S \to s$ now have actual values (data) they are no longer random

    - No way to know if parameters are really contained in interval (looking into past).
    - This is evidence for statement of probability, one does not really know if this true (high level of belief)

Nhớ $\bar{X}$, S random variables, what will happen, before experiment

BIVARIATE NORMAL $N\left[\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix}, \begin{bmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_y^2\end{bmatrix}\right]$ with $\sigma_{xy} = \rho\sigma_x\sigma_y$ and $\sigma_x^2 \equiv \sigma_{xx}$

Let U and V have joint density

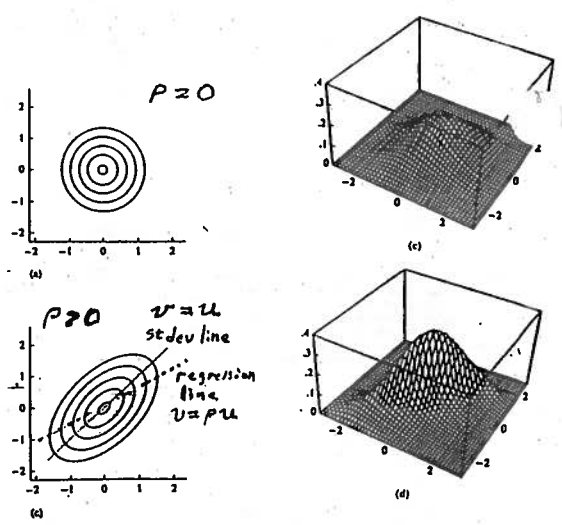
(1) $$f_{U,V}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{v^2 - 2\rho uv + u^2\}}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{(v-\rho u)^2 + (1-\rho^2)u^2\}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

$$= f_{V|U=u}(v) \cdot f_U(u)$$

(2) $$= f_{N(\rho u,\, 1-\rho^2)}(v) \cdot f_{N(0,1)}(u) .$$



Now let $U \equiv \frac{X-\mu_X}{\sigma_X}$ and $V \equiv \frac{Y-\mu_Y}{\sigma_Y}$.

Then

(3) $$f_{X,Y}(x,y) = \frac{1}{\sigma_X\sigma_Y} f_{U,V}\left(\frac{x-\mu_X}{\sigma_X}, \frac{y-\mu_Y}{\sigma_Y}\right) = \left(\begin{matrix}\text{book}\\ \text{formula}\end{matrix}\right)$$

$$= \frac{1}{\sigma_Y} f_{V|U=u}(v) \cdot \frac{1}{\sigma_X} f_U(u)$$

(4) $$= f_{Y|X=x}(y) \cdot f_X(x)$$

where

(5) $$f_X(x) = \frac{1}{\sigma_X} f_U(u) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \text{ is } N(\mu_X, \sigma_X^2)$$

and

$$f_{Y|X=x}(y) = \frac{1}{\sigma_Y} f_{V|U=u}(v)$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_Y^2(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_Y}{\sigma_Y} - \rho\frac{\sigma_Y}{\sigma_Y}\frac{x-\mu_X}{\sigma_X}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_\epsilon} e^{-\frac{1}{2\sigma_\epsilon^2}(y - m(x))^2}$$

(6) $$= f_{N(m(x),\, \sigma_\epsilon^2)}(y)$$

with

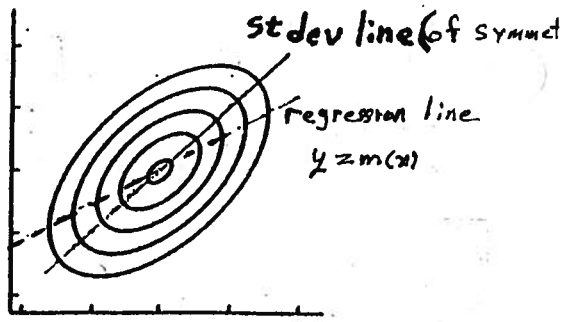
(7) $$m(x) \equiv \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$

$$= \left(\mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X\right) + \left(\rho\frac{\sigma_Y}{\sigma_X}\right)x$$

(8) $$\equiv \alpha + \beta x$$

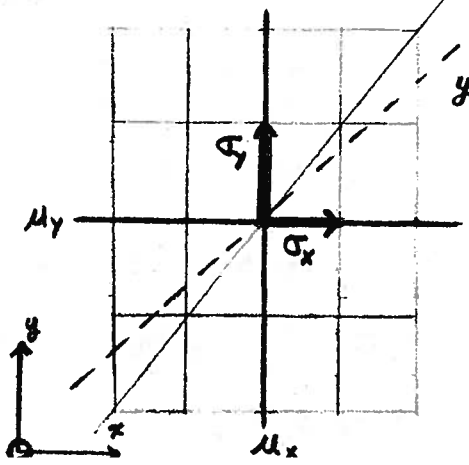
and

(9) $$\sigma_\epsilon^2 \equiv \sigma_Y^2(1-\rho^2) = v^2(x) .$$



So

$Y|X=x$ is $N(m(x), \sigma_\epsilon^2)$



stdev line: $y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(x-\mu_X)$

regression line

$y = m(x)$

$= \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$

$= \alpha + \beta x$

Assuming (X,Y) is Bivariate Normal implies that

$Y|X=x$ is $N(m(x), \sigma_\epsilon^2)$

The y intercept and slope for a line

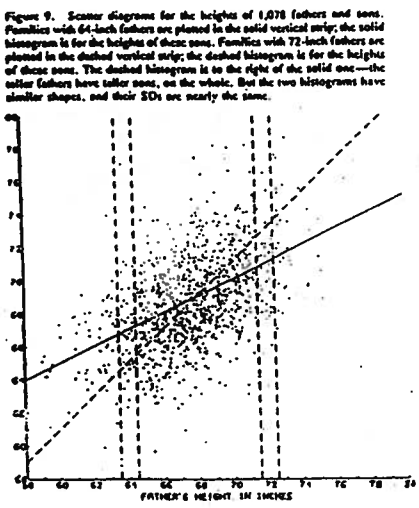
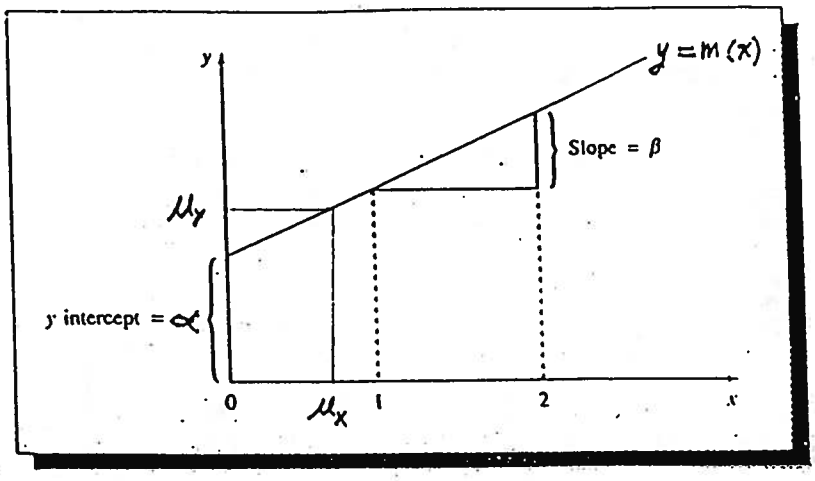


Figure 9. Scatter diagrams for the heights of 1,078 fathers and sons. Families with 64-inch fathers are plotted in the solid vertical strip; the solid histogram is for the heights of these sons. Families with 72-inch fathers are plotted in the dashed vertical strip; the dashed histogram is for the heights of these sons. The dashed histogram is to the right of the solid one—the taller fathers have taller sons, on the whole. But the two histograms have similar shapes, and their SDs are nearly the same.

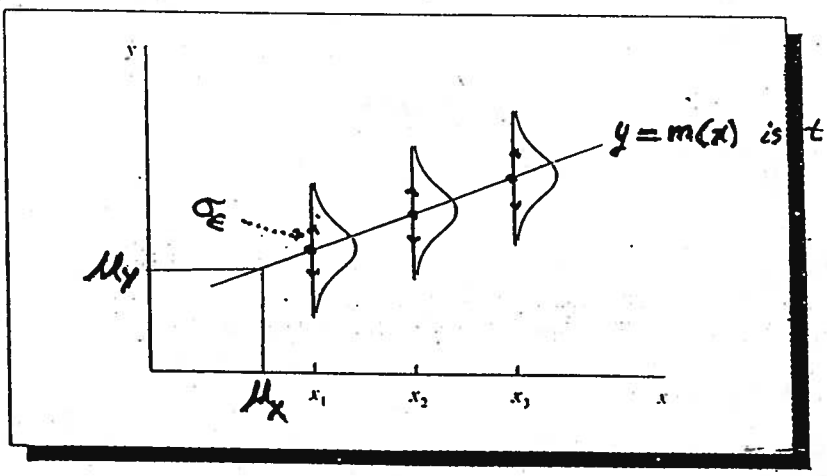


The linear model

$$m(x) \equiv E(Y|X=x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x-\mu_X) \equiv \mu_Y + \beta(x-\mu_X)$$
$$= (\mu_Y - \beta\mu_X) + \beta x \equiv \alpha + \beta x$$

So, alternatively:

Linear probabilistic model



$Y_i = m(x_i) + \epsilon_i$ for $1 \le i \le n$

where $\epsilon_1, \ldots, \epsilon_n$ are independent $N(0, \sigma_\epsilon^2)$

and $m(x) = \alpha + \beta x = \mu_Y + \beta(x-\mu_X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$

This last is called the simple linear regression model.

Note: $v^2(x) \equiv Var[Y|X=x] = \sigma_Y^2(1-\rho^2) \equiv \sigma_\epsilon^2$.

Basic regression model

$Y = \gamma + \beta x + \epsilon$, with $\epsilon = N(0, \sigma_\epsilon^2)$

$m(x) = (\gamma + \beta\bar{x}) + \beta(x_i - \bar{x})$

- mean value of X, m(x) should depend on X (assumed linear)
- ϵ errors or random fluctuations, inherent variation (assumed const)
- γ height at origin
- β slope

Derivation of $\mathcal{T}_n$ and $\mathcal{F}_{m,n}$

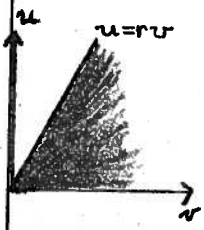
Def 1. Let $Z \cong N(0,1)$, $U \cong \chi^2_m$, and $V \cong \chi^2_n$ be independent. Define

(1) $$T \equiv \frac{Z}{\sqrt{\frac{1}{n}V}} \cong \mathcal{T}_n \quad \text{and} \quad W \equiv \frac{\frac{1}{m}U}{\frac{1}{n}V} \cong \mathcal{F}_{m,n}$$

to have the **Student** $\mathcal{T}_n$ and **Snedecor** $\mathcal{F}_{m,n}$ distributions, respectively.

Theorem 1. $W \cong \mathcal{F}_{m,n}$ where this is the Snedecor $\mathcal{F}_{m,n}$-distribution with

(2) $$f_W(w) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \frac{m^{m/2} n^{n/2} w^{\frac{m}{2}-1}}{(n+mw)^{(m+n)/2}} \quad \text{on } [0,\infty).$$



Proof. Let $R \equiv U/V$ for U and V independent with densities on $[0,\infty)$. Then

$$\int_0^r f_R(w)\,dw = F_R(r) = P(U/V \le r) = P(U \le rV)$$
$$= \int_0^\infty \int_0^{rv} f_U(u) f_V(v)\,du\,dv$$
$$= \int_0^\infty \int_0^r f_U(vw) f_V(v)\, v\,dw\,dv \quad \text{letting } w \equiv u/v \text{ with } du = v\,dv$$
$$= \int_0^r \underbrace{\left[\int_0^\infty f_U(vw) f_V(v)\, v\,dv \right]}_{= f_R(w)} dw .$$

(3) ← useful formula, in general.

So, when $U \cong \chi^2_m \cong \text{Gamma}(r = \frac{m}{2}, \lambda = \frac{1}{2})$ and $V \cong \chi^2_n \cong \text{Gamma}(r = \frac{n}{2}, \lambda = \frac{1}{2})$, we have

(a) $$f_R(r) = \int_0^\infty f_U(vr) f_V(v)\,dv = \int_0^\infty \frac{(rv)^{\frac{m}{2}-1}}{2^{m/2}\Gamma(\frac{m}{2})} e^{-\frac{rv}{2}} \cdot \frac{v^{\frac{n}{2}-1}}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{v}{2}} \cdot v\,dv$$

(b) $$= \frac{1}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2}) 2^{(m+n)/2}} r^{\frac{m}{2}-1} \cdot \text{stuff} \cdot \underbrace{\frac{1}{\text{stuff}} \int_0^\infty v^{\frac{m+n}{2}-1} e^{-\frac{1}{2}(r+1)v}\,dv}_{=1}$$

(c) $$= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \frac{r^{\frac{m}{2}-1}}{(1+r)^{(m+n)/2}} \quad \text{on } 0 \le r < \infty .$$

Since $W \equiv \frac{n}{m} R$, we have $F_W(w) = P(W \le r) = P(\frac{n}{m} R \le w) = P(R \le \frac{m}{n} w) = F_R(\frac{m}{n} w)$; so

(d) $$f_W(w) = f_R(\tfrac{m}{n} w) \cdot \tfrac{m}{n} = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \frac{(\frac{mw}{n})^{\frac{m}{2}-1}}{(1+\frac{mw}{n})^{(m+n)/2}} \cdot \frac{m}{n} = (\text{expression (2)}).$$ ∎

Theorem 2. $T \cong \mathcal{T}_n$ where this is the Student $\mathcal{T}_n$-distribution with

(4) $$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})} \frac{1}{(1+\frac{t^2}{n})^{(n+1)/2}} \quad \text{for } -\infty < t < \infty .$$

Proof. For $t > 0$, $F_T(t) = P(T \le t) = \frac{1}{2} + P(0 \le T \le t) = \frac{1}{2} + \frac{1}{2} P(-t \le T \le t) = \frac{1}{2} + \frac{1}{2} F_{T^2}(t^2)$, so

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{d}{dt}\left[\tfrac{1}{2} + \tfrac{1}{2} F_{T^2}(t^2)\right] = \tfrac{1}{2} f_{T^2}(t^2) \cdot 2t = t\, f_{T^2}(t^2)$$

(e) $$= t\, f_W(t^2) \quad \text{since } T^2 = \left(\frac{Z}{\sqrt{\frac{1}{n}V}}\right)^2 = \frac{\frac{1}{1}Z^2}{\frac{1}{n}V} \cong \mathcal{F}_{1,n} \equiv W$$

and $f_W(\cdot)$ comes from expression (2) with $m = 1$

(f) $$= \frac{t\,\Gamma(\frac{1+n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} \frac{1^{1/2} n^{n/2} (t^2)^{(\frac{1}{2}-1)}}{(n+t^2)^{(1+n)/2}} = (\text{expression (4)}).$$ ∎

Two sample problem

Example 1 Let $X_1, \dots, X_n$ be ind $N(\mu, \sigma^2)$. We know that

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \equiv N(0,1) \quad \text{and} \quad \frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} = \frac{SS_{tot}}{\sigma^2} = \chi^2_{n-1}$$

(1) are independent

Thus

$$(2) \quad T \equiv \frac{\sqrt{n}(\bar{X} - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{\frac{SS_{xx}}{n-1}/\sigma^2}} \cong \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \equiv T_{n-1}$$

Example 2: Let $X_1, \dots, X_n$ be iid $N(\mu, \sigma^2)$ and $Y_1, \dots, Y_n$ be iid

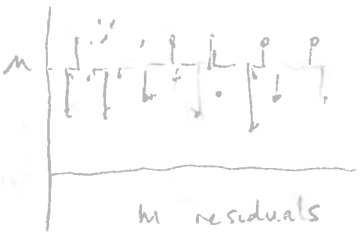
$Y_1, \dots, Y_n$ ind $N(\nu, \sigma^2)$

$\theta = \mu - \nu$ is estimated by $\hat{\theta} \equiv \hat{\theta}_{m,n} = \bar{X}_m - \bar{Y}_n$

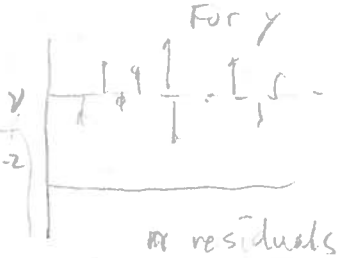
$$\hat{\theta} \equiv N\left(\mu - \nu, \frac{\sigma^2}{m} + \frac{\sigma^2}{n} = \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right) = \sigma^2\left(\frac{m+n}{mn}\right)\right)$$

$$(3) \quad Z = Z_{mn} = \frac{(\bar{X} - \bar{Y}) - (\mu - \nu)}{\sigma\sqrt{\frac{m+n}{mn}}} = \sqrt{\frac{mn}{m+n}}\,\frac{(\bar{X} - \bar{Y}) - (\mu - \nu)}{\sigma}$$

For X



For Y



$$SS_{xx} \equiv \sigma^2 \chi^2_{m-1}$$

$$S_{xy} \equiv \sigma^2 \chi^2_{n-1}$$

$$SS_{xx} + S_{yy} = \sigma^2 \chi^2_{m+n-2}$$

So $S_p^2 \equiv \frac{SS_{xx} + S_{yy}}{n+m-2} = \sigma^2 \frac{\chi^2_{m+n-2}}{m+n-2}$ — S_p = S pooled

$$\therefore T \equiv T_{mn} \equiv \sqrt{\frac{mn}{m+n}}\,\frac{(\bar{X} - \bar{Y}) - (\mu - \nu)/\sigma}{\sqrt{S_p^2/\sigma^2}} = \frac{Z}{\sqrt{\frac{\chi^2_{m+n-2}}{m+n-2}}} = \tilde{t}_{m+n-2}$$

Confidence interval for $\mu - \nu$:

$$(\bar{X} - \bar{Y}) \pm t\sqrt{\frac{m+n}{mn}}\,S_p$$

$$(4) \quad S^2 \equiv S_p^2 \equiv \frac{(m-1)S_x^2 + (n-1)S_y^2}{(m+n-2)}$$

$$S_x^2 = \frac{SS_{xx}}{m-1} \qquad S_y^2 = \frac{SS_{xx}}{n-1}$$

Also, $\frac{S_x^2}{S_y^2} = \frac{SS_{xx}/(m-1)}{SS_{yy}/(n-1)} \cong \tilde{f}_{m-1,n-1}$

even if the two variances are different

$$1 - \alpha = P\left(f_L \le \frac{\sigma_y^2}{\sigma_x^2}\frac{S_x^2}{S_y^2} \le f_u\right)$$

$$= P\left(\frac{S_x^2}{S_y^2} \cdot f_L \le \frac{\sigma_x^2}{\sigma_y^2} \le \frac{S_x^2}{S_y^2} \cdot f_u\right)$$

Moments Via Matrices

Let

$$(1)\quad \underline{Y} \equiv \begin{bmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1n}X_n \\ \vdots \\ a_{i1}X_1 + \cdots + a_{in}X_n \\ \vdots \\ a_{n1}X_1 + \cdots + a_{nn}X_n \end{bmatrix} \equiv A\underline{X} = \begin{bmatrix} a_{11} & \cdots & a_{nn} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \underline{X} = \begin{bmatrix} \underline{a}_1' \\ \vdots \\ \underline{a}_i' \\ \vdots \\ \underline{a}_n' \end{bmatrix} \underline{X} \quad \text{and} \quad \underline{\mu}_X = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

with $\mu_i \equiv EX_i$.

Then

$$EY_i = E(a_{i1}X_1 + \cdots a_{in}X_n) = a_{i1}\mu_1 + \cdots + a_{in}\mu_n = \underline{a}_i'\underline{\mu},$$

so

$$(2)\quad \underline{\mu}_Y = A\underline{\mu}_X .$$

Also,

$$\Sigma_Y \equiv [\sigma_{ij}] \equiv \left[[\mathrm{Cov}[Y_i, Y_j]]\right] = \left[\mathrm{Cov}\left[\sum_{k=1}^{n} a_{ik}X_k, \sum_{l=1}^{n} a_{jl}X_l\right]\right]$$

$$= \left[\sum_{l=1}^{n}\sum_{k=1}^{n} a_{ik}\,\sigma_{X\,kl}\,a_{jl}\right] = \left[\underline{a}_i'\,\Sigma_X\,\underline{a}_j\right]$$

$$(3)\quad = A\Sigma_X A' .$$

Summary

$$(4)\quad \underline{Y} \equiv A\underline{X} + \underline{c} \cong (A\underline{\mu}_X + \underline{c},\ A\Sigma_X A') = (\mu_Y, \Sigma_Y)$$

Additionally, if

$$\underline{U}^{m\times 1} = A^{m\times n}\,\underline{X}^{n\times 1}$$

$$\underline{V}^{I\times 1} = B^{I\times J}\,\underline{Y}^{J\times 1}$$

on some (Σ, A, P),

then

$$\Sigma_{\underline{U},\underline{V}}^{m\times I} = \left[\mathrm{Cov}[U_\alpha, V_\beta]\right] = \left[\mathrm{Cov}\left[\sum_{\gamma=1}^{n} a_{\alpha\gamma}X_\gamma, \sum_{\delta=1}^{J} b_{\beta\delta}Y_\delta\right]\right]$$

$$= \left[\sum_{\gamma=1}^{n} a_{\alpha\gamma}\sum_{\delta=1}^{J} b_{\beta\delta}\,\mathrm{Cov}[X_\gamma, Y_\delta]\right]$$

$$= A^{m\times n}\,\Sigma_{X,Y}^{n\times J}\,B'^{\,J\times I} = A\Sigma_{X,Y}B' .$$

EXAMPLE 2. (a) Let $X \cong (3,5)$ and define $Y = 2X+3$. Then

$$\mu_Y = 2\cdot 3 + 3 = 9 \quad \text{and} \quad \sigma_Y^2 = 2^2 \sigma_X^2 = 20 .$$

(b) Suppose $X \cong (3,5)$ and $Y \cong (1,4)$ with $\sigma_{XY} = 3$. Let

$$U \equiv 2X - Y \quad \text{and} \quad V \equiv X + 3Y .$$

The moments of these new rvs are:

$$\mu_U = 2\mu_X - \mu_Y = 5 \quad \text{and} \quad \mu_V = 3 + 3\cdot 1 = 6 .$$

$$\sigma_U^2 = 4\sigma_X^2 + \sigma_Y^2 - 4\sigma_{XY} = 4\cdot 5 + 4 - 4\cdot 3 = 12 .$$

$$\sigma_V^2 = \sigma_X^2 + 9\sigma_Y^2 + 6\sigma_{XY} = 5 + 9\cdot 4 + 6\cdot 3 = 59 .$$

$$\sigma_{UV} = 2\sigma_X^2 + 6\sigma_{XY} - \sigma_{XY} - 3\sigma_Y^2 = 2\cdot 5 + 5\cdot 3 - 3\cdot 4 = 13 .$$

Thus correlations are given by

$$\rho_{UV} = \frac{\sigma_{UV}}{\sigma_U \sigma_V} = \frac{13}{\sqrt{12\cdot 59}} = .4886 \quad \text{and} \quad \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{3}{\sqrt{5\cdot 4}} = .6708 .$$

EXAMPLE 3. Let $X \cong (0,1)$ and $Y \cong (0,1)$ be uncorrelated.

(a) Then $X+Y$ and $X-Y$ are uncorrelated rvs, and both are $(0,2)$.

(b) Also $\frac{X+Y}{\sqrt{2}}$ and $\frac{X-Y}{\sqrt{2}}$ are uncorrelated rvs, and both are $(0,1)$.

HW **EXERCISE 1** Suppose $X \cong (2,6)$ and $Y \cong (-1,9)$ with $\sigma_{XY} = -2$. Let

$$U = 3X + 2Y \quad \text{and} \quad V = 2X - 3Y .$$

Determine the means, variances, covariance, and correlation for U and V.

EXAMPLE 4 Draw a card from a well shuffled deck.

Let X equal 1 or 0 as the card is a "heart" or not.

Let Y equal 1 or 0 as the card is a "club" or not.

Let Z equal 1 or 0 as the card is a "King" or not. Then:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E(XY) - \mu_X \mu_Y}{\sigma_X \sigma_Y} = \frac{0 - \frac{1}{4}\cdot\frac{1}{4}}{\sqrt{\frac{1}{4}\cdot\frac{3}{4}}\sqrt{\frac{1}{4}\cdot\frac{3}{4}}} = \frac{-\frac{1}{16}}{\frac{3}{16}} = -\frac{1}{3}.$$

$$\rho_{XZ} = \frac{\sigma_{XZ}}{\sigma_X \sigma_Z} = \frac{E(XZ) - \mu_X \mu_Z}{\sigma_X \sigma_Z} = \frac{\frac{1}{52} - \frac{1}{4}\cdot\frac{1}{13}}{\sqrt{\frac{1}{4}\cdot\frac{3}{4}}\sqrt{\frac{1}{13}\cdot\frac{12}{13}}} = \frac{0}{\frac{6}{2\cdot 13}} = 0.$$

HW **EXERCISE 2** Roll a die.

Let X equal 1 or 0 as the number is a prime or not.

Let Y equal 1 or 0 as the number is a perfect square or not.

Let $U \equiv X + 2Y$. Evaluate μ_U and σ_U^2.

EXAMPLE 5 Let I be chosen at random from 1, 2, 3, 4, 5, 6.

Let $X_i \cong \text{Binomial}(i, \frac{1}{2})$ for $1 \le i \le 6$. Let $Y \equiv X_I$. Then:

$$\mu_Y = E(E(Y|I)) = \sum_{i=1}^{6} E(Y|I=i)P(I=i) = \frac{1}{6}\sum_{i=1}^{6}\frac{i}{2} = \frac{6\cdot 7}{24} = \frac{7}{4}.$$

$$EY^2 = E(E(Y^2|I)) = \sum_{i=1}^{6} E(Y^2|I=i)P(I=i) = \frac{1}{6}\sum_{i=1}^{6}\left[\frac{i}{4} + \left(\frac{i}{2}\right)^2\right] = \frac{6\cdot 7}{24\cdot 2} + \frac{6\cdot 7\cdot 13}{24\cdot 6} = \frac{14}{3}.$$

$$\sigma_Y^2 = EY^2 - \mu_Y^2 = \frac{14}{3} - \left(\frac{7}{4}\right)^2 = \frac{77}{48} = 1.6042.$$

Note also that:

$$E(IY) = E(I\,E(Y|I)) = \frac{1}{6}\sum_{i=1}^{6} i\,E(Y|I=i) = \frac{1}{6}\sum_{i=1}^{6} i\,\frac{i}{2} = \frac{6\cdot 7\cdot 13}{12\cdot 6} = \frac{91}{12}.$$

$$\text{Cov}[I,Y] = \frac{91}{12} - \frac{7}{2}\cdot\frac{7}{4} = \frac{35}{24} \quad \text{and} \quad \rho_{I,Y} = \quad .$$

EXERCISE 3. ( ) The number of insurance claims N_t filed between times 0 and t is modeled as a Poisson process with intensity ν. Suppose that the first claim, the second claim, the third claim, ... filed are for $\overset{iid}{}$ dollar amounts $U_1, U_2, U_3, \ldots$ with μ_U and σ_U^2. Thus the total dollar value T_t of all claims filed by time t is $T_t = \sum_{i=1}^{N(t)} U$.

Write out the complete details of the calculation evaluating $\mu_t \equiv ET_t$ and $\sigma_t^2 \equiv \text{Var}[T_t]$.

EXAMPLE 6 Each of N people put their name on a ball and place it in an urn. The urn is shaken and each person draws a ball with a name. Let T_N denote the number of people who draw their own name. Then

$$\mu_N \equiv E T_N = 1 \quad \text{and} \quad \sigma_N^2 \equiv \text{Var}[T_N] = 1 .$$

(Note that T_N is the sum of N dependent Bernoulli$(\frac{1}{N})$ rvs. Perhaps, also, $T_N \to_d$ Poisson(1), since Poisson(1) $\cong (1,1)$ also.) Well

$$T_N = X_1 + \cdots + X_N \quad \text{where } X_i \text{ equals 1 or 0 as "person } i\text{" draws their own hat, or not.}$$

So

$$\mu_N = EX_1 + \cdots + EX_N = N\,EX_1 = N\left\{ 1 \cdot \tfrac{1}{N} + 0 \cdot \tfrac{N-1}{N} \right\} = 1$$

and

$$\sigma_N^2 = N\,\text{Var}[X_1] + N(N-1)\,\text{Cov}[X_1, X_2]$$

$$= N\left\{ \tfrac{1}{N}\left(1 - \tfrac{1}{N}\right) \right\} + N(N-1)\left\{ \tfrac{1}{N} \cdot \tfrac{1}{N-1} - \tfrac{1}{N} \cdot \tfrac{1}{N} \right\} = \left(1 - \tfrac{1}{N}\right) + N(N-1)\tfrac{1}{N^2(N-1)} = 1 .$$

Note that

$$\text{Corr}[X_1, X_2] = \frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{\frac{1}{N^2(N-1)}}{\frac{N-1}{N^2}} = \frac{1}{(N-1)^2} = (\text{tiny}) .$$

HW **EXERCISE 4** Choose two cards at random from a well shuffled deck. Let I_1 (or I_2) be 1 or 0 as the first (as the second) card drawn is a Queen. Let K_1 (or K_2) be 1 or 0 as the first (as the second) card drawn is a King. Let $M = I_1 + I_2$ and $N \equiv K_1 + K_2$ denote the number of Queens and Kings in our hand of size two.

(a) Compute EI_1, $\text{Var}[I_1]$, and $\text{Cov}[I_1, I_1]$.

(b) Compute $\text{Cov}[I_1, K_1]$ and $\text{Cov}[I_1, K_2]$.

(c) Compute $\text{Var}[M]$ and $\text{Var}[N]$.

EXAMPLE 7. Let $X_1, \ldots, X_n$ be iid (μ, σ^2) rvs. Then:

$$\mathrm{Cov}[X_1, \bar{X}] = \mathrm{Cov}\left[X_1, \tfrac{1}{n}\textstyle\sum_1^n X_i\right] = \tfrac{1}{n}\textstyle\sum_1^n \mathrm{Cov}[X_1, X_i] = \tfrac{1}{n}[\sigma^2 + 0 + \cdots + 0] = \tfrac{\sigma^2}{n}.$$

or

$$\mathrm{Cov}[X_2, \bar{X}] = \mathrm{Cov}\left[X_1, \tfrac{1}{n}(X_1 + \cdots + X_n)\right] = \tfrac{1}{n}\mathrm{Cov}[X_1, X_1 + \cdots + X_n] = \tfrac{1}{n}[\sigma^2 + 0 + \cdots + 0] = \tfrac{\sigma^2}{n}.$$

Also,

$$\begin{aligned}\mathrm{Var}[X_1 - \bar{X}] &= \mathrm{Cov}[X_1 - \bar{X}, X_1 - \bar{X}] = \mathrm{Cov}[X_1, X_1] - \mathrm{Cov}[X_1, \bar{X}] - \mathrm{Cov}[X_1, \bar{X}] + \mathrm{Cov}[\bar{X}, \bar{X}] \\ &= \mathrm{Var}[X_1] - 2\,\mathrm{Cov}[X_1, \bar{X}] + \mathrm{Cov}[\bar{X}, \bar{X}] = \sigma^2 - 2\tfrac{\sigma^2}{n} + \tfrac{\sigma^2}{n} = \sigma^2\left(1 - \tfrac{1}{n}\right) \\ &= \tfrac{n-1}{n}\sigma^2,\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}[X_1 - \bar{X}, X_2 - \bar{X}] &= \mathrm{Cov}[X_1, X_2] - \mathrm{Cov}[X_1, \bar{X}] - \mathrm{Cov}[X_2, \bar{X}] + \mathrm{Cov}[\bar{X}, \bar{X}] \\ &= 0 - \tfrac{\sigma^2}{n} - \tfrac{\sigma^2}{n} + \tfrac{\sigma^2}{n} = -\tfrac{\sigma^2}{n}.\end{aligned}$$

Thus

$$\mathrm{Corr}[X_1 - \bar{X}, X_2 - \bar{X}] = \frac{\mathrm{Cov}[X_1 - \bar{X}, X_2 - \bar{X}]}{\sqrt{\mathrm{Var}[X_1 - \bar{X}]\ \mathrm{Var}[X_2 - \bar{X}]}} = \frac{-\frac{\sigma^2}{n}}{\sqrt{\frac{n-1}{n}\sigma^2 \cdot \frac{n-1}{n}\sigma^2}} = -\frac{1}{n-1}.$$

EXERCISE 5 Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from (μ, σ^2) and (ν, τ^2), respectively. Let $\bar{Z} \equiv \frac{X_1 + \cdots + X_m + Y_1 + \cdots + Y_n}{m+n}$.

(a) Evaluate $\mathrm{Cov}[\bar{X} - \bar{Z}, \bar{Y} - \bar{Z}]$ and $\mathrm{Corr}[\bar{X} - \bar{Z}, \bar{Y} - \bar{Z}]$.

(b) Simplify these expressions when $\sigma^2 = \tau^2$.

(c) Evaluate $\mathrm{Cov}[X_i - \bar{Z}, \bar{X} - \bar{Y}]$, and simplify it when $\sigma^2 = \tau^2$.

EXAMPLE 8. Twenty people (comprising 10 married couples) are randomly seated around a table. Let N denote the number of wives seated next to their husbands. Compute μ_N and σ_N^2. Well, $N = \sum_{i=1}^{10} X_i$ where X_i equals 1 if "Mrs i" is next to "Mr i", and 0 if not. Then:

$$\mu_N = 10\,E X_1 = 10\,P(X_1 = 1) = 10 \cdot \tfrac{2}{19} = \tfrac{20}{19}.$$

$$\sigma_N^2 = \mathrm{Var}\left[\sum_{1}^{10} X_i\right] = \mathrm{Cov}\left[\sum_{1}^{10} X_i, \sum_{1}^{10} X_j\right] = 10\,\mathrm{Var}[X_1] + 10 \cdot 9\,\mathrm{Cov}[X_1, X_2]$$

$$= 10\left\{\tfrac{2}{19}\left(1 - \tfrac{2}{19}\right)\right] + 90\left\{E(X_1 X_2) - \tfrac{2}{19} \cdot \tfrac{2}{19}\right\}$$

$$= 10\left\{\tfrac{2}{19} \cdot \tfrac{17}{19}\right\} + 90\left\{1 \cdot \tfrac{2}{19} \cdot \left(\tfrac{2}{18} \cdot \tfrac{1}{17} + \tfrac{16}{18} \cdot \tfrac{2}{17}\right) - \tfrac{2}{19} \cdot \tfrac{2}{19}\right\} = \cdots = \tfrac{360}{361}.$$

(Perhaps this also converges to a Poisson(1) as the number of couples increases.) □

HW **EXERCISE C.** An urn contains 5 Reds, 10 Whites, and 5 Blues. We draw 10 pairs from the urn, and let N denote the number of {Red, Blue} pairs. Compute μ_N and σ_N^2 (and do it two ways).

5 R
10 W
5 B

Method I: $N = \sum_{i=1}^{5} X_i$ where X_i equals 1 if "serial number i" of Red is paired with a "Blue", and is 0 otherwise.

Method II: $N = \sum_{j=1}^{10} Y_j$ where Y_j equals 1 if the j-th pair drawn is {Red, Blue}, and is 0 otherwise.

EXAMPLE 9. (Multinomial)

Partition the sample space $A_1 + \cdots + A_I = S$

with probabilities $p_1 + \cdots + p_I = 1$

Let $\vec{X} \equiv (X_1, \ldots, X_I)'$

where $X_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{else} \end{cases}$ So, $\vec{X} \cong \text{Multinomial}(n=1, p_1, \ldots p_I)$.

Now let $\vec{X}_1 \equiv (X_{11}, \ldots, X_{I1})'$

$\vdots$

$\vec{X}_n \equiv (X_{1n}, \ldots, X_{In})'$

denote n iid repetitions.

Let $\vec{N} \equiv (N_1, \ldots, N_I)'$

where $N_i \equiv \sum_{k=1}^{n} X_{ik}$ denote the number of times that A_i occurs.

Then $\vec{N} = (N_1, \ldots, N_I)' = \sum_{k=1}^{n} \vec{X}_k' \cong \text{Multinomial}(n; \vec{p}) = \text{Multinomial}(n; p_1, \ldots, p_I)$;

so $P(\vec{N} = \vec{m}) = \binom{n}{m_1, \ldots, m_I} p_1^{m_1} \cdots p_I^{m_I}$ for ≥ 0 integers $m_1 + \cdots + m_I = n$.

Consider the basic $\vec{X}$-experiment (with $\vec{X} \equiv (X_1, \ldots, X_I)$, as above).

Trivially, $E X_i = p_i$, $\text{Var}[X_i] = p_i(1-p_i)$, $E(X_i X_j) = 0$, $\text{Cov}[X_i, X_j] = -p_i p_j$.

Thus $\mu_{\vec{X}} = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_I} \end{bmatrix} = \vec{p}$

and

$$\Sigma_{\vec{X}} \equiv \left[\!\left[\sigma_{ij} = \text{Cov}[X_i, X_j] \right]\!\right] = \left[\!\left[\begin{matrix} p_1(1-p_1) & & -p_i p_j \\ & \ddots & \\ -p_i p_j & & p_I(1-p_I) \end{matrix} \right]\!\right]$$

$$= \begin{bmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_I \end{bmatrix} - \begin{bmatrix} p_1 p_1 & & p_i p_j \\ & \ddots & \\ p_i p_j & & p_I p_I \end{bmatrix}$$

$$= D_{\vec{p}} - \vec{p}\vec{p}'$$

Since $\vec{N} \equiv \sum_{k=1}^{n} \vec{X}_k$, we thus have $\vec{N} \cong (n\vec{p}, n[D_{\vec{p}} - \vec{p}\vec{p}'])$.

Comment (Multivariate CLT) $\sqrt{n}(\vec{\bar{X}} - \vec{\mu}) \to_d N(0, \Sigma_{\vec{X}})$ when $\vec{X}_1, \ldots, \vec{X}_n$ are iid $(\vec{\mu}, \Sigma_{\vec{X}})$.

In the present case, let $\vec{\hat{p}} \equiv \frac{\vec{N}}{n}$ and obtain $\sqrt{n}(\vec{\hat{p}} - \vec{p}) \to_d N(\vec{0}, D_{\vec{p}} - \vec{p}\vec{p}')$. □

Exercise Sheet

Exercises—B

1. For Example A2, show that Q is the MLE for θ. What is the MLE for θ^3? $X_i \cong N(i\theta, 1)$ and ind.
2. For Example A3, show that R is the MLE for θ. What is the MLE for θ^2? $X_i \cong Exp(mean = i\theta)$ and ind.
3. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim P(\theta)$. What is the MLE of θ? of e^{θ}?
4. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim \Gamma(i, \theta)$. What are the MLEs of θ and θ^{-1}?
5. Let $X_1, \ldots, X_n$ be independent, with X_i geometrically distributed with parameter θ. That is, X_i has density $f(x;\theta) = \theta(1-\theta)^x, x = 0, 1, \ldots, 0 < \theta < 1$. Find the MLEs of θ and $(1-\theta)^{-1}$.
6. Let $X_1, \ldots, X_n$ be independently identically distributed with X_i having density $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$. Find the MLE of θ.
7. For Example C in Section 7.1.1, find the MLE of (θ, σ). What is the MLE of $\theta\sigma$? of $\theta\sigma^2$? $X_i \cong N(i\theta, \sigma^2)$
8. Let (X, Y) have a trinomial distribution with parameters n, θ^2, $2\theta(1-\theta)$, and $(1-\theta)^2$. That is, (X, Y) has joint density function

   $$f(x,y) = k(x,y)(\theta^2)^x(2\theta(1-\theta))^y((1-\theta)^2)^{n-x-y}, \quad x = 0, 1, \ldots, y = 0, 1, \ldots, x + y \leq n.$$

   Suppose n is known. Show that the MLE of θ is $(2X + Y)/2n$.
9. Let (X, Y) have a trinomial distribution with parameters n, θ_1, and θ_2. Show that the MLE of (θ_1, θ_2) is $(X/n, Y/n)$.

Exercises—C

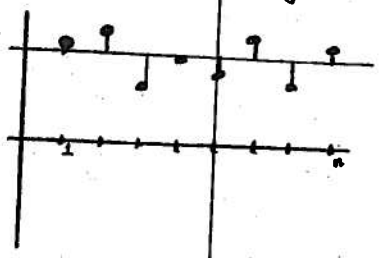
1. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(\theta, \theta^2)$. Find the MLE of θ. (Be sure to verify which root of the quadratic leads to the maximum.)
2. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(\mu, \delta)$. That is, X_i is normally distributed with mean μ and variance δ. Hence, $\theta = (\mu, \delta)$ for this model. Without using the invariance principle, show that the MLE for (μ, δ) is given by $\hat{\mu} = \bar{X}, \hat{\delta} = (n-1)S^2/n$.

Review of Classical (Intuitive) Statistics

7-MAY-2004

Example 1. (One sample problem) Suppose

(1) $Y_i = \mu + \epsilon_i$ (for $1 \le i \le n$) with $\epsilon_1, \ldots, \epsilon_n$ ind $N(0, \sigma^2)$.



The obvious estimator of μ is $\bar{Y}$.

The least squares estimator (LSE) of μ is $\bar{Y}$.

The maximum likelihood estimator (MLE) of μ is $\bar{Y}$.

How good an estimator is it? The <u>power lies in the residuals</u>

(2) $\hat{\epsilon}_i \equiv Y_i - \bar{Y}$ = observation$_i$ − fitted value$_i$ ≡ residual$_i$.

These <u>allow us to</u> estimate σ, and thus <u>quantify</u> the quality of $\bar{Y}$ as an estimator of μ. Thus (<u>under normality</u>)

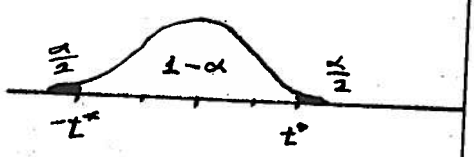
(3) $SS_{YY} \equiv \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \cong \sigma^2 \chi^2_{n-1} \cong \sigma^2 \times (df, 2df) \cong \sigma^2 (n-1, 2(n-1))$

Hence

(4) $S^2 \equiv \frac{SS_{YY}}{n-1} \cong \sigma^2 \frac{\chi^2_{n-1}}{n-1} \cong \sigma^2 \times \left(1, \frac{2}{n-1}\right) \cong \left(\sigma^2, \frac{2\sigma^4}{n-1}\right)$,

showing that S^2 is both unbiased and consistent. Since

(5) $T \equiv \frac{\sqrt{n}(\bar{Y} - \mu)}{S} \cong \mathcal{T}_{n-1}$,



we have

$$1 - \alpha = P(-t^* \le T \le t^*) = P\left(-t^* \le \frac{\sqrt{n}(\bar{Y} - \mu)}{S} \le t^*\right)$$

(6) $$= P\left(\bar{Y} - t^* \frac{S}{\sqrt{n}} \le \mu \le \bar{Y} + t^* \frac{S}{\sqrt{n}}\right) = P\left(\mu \in \bar{Y} \pm t^* \frac{S}{\sqrt{n}}\right),$$

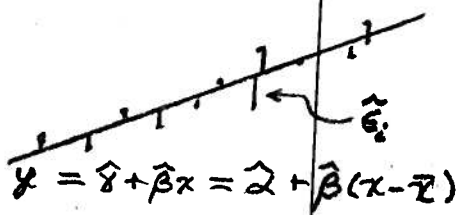
giving the $1-\alpha$ confidence interval $\bar{Y} \pm t^* \frac{S}{\sqrt{n}}$ for the fixed but unknown μ.

[Being the LSE meant that $a = \bar{Y}$ minimizes the lack of fit $SS(a) \equiv \sum_1^n (Y_i - a)^2$.]

Example 2. (Two Sample Problem)

Example 3. (Simple linear regression) Let

$$Y_i = \gamma + \beta x_i + \epsilon_i \equiv \alpha + \beta(x_i - \bar{x}) + \epsilon_i$$



Let $\hat{\alpha} + \hat{\beta}$ minimize $SS(a,b) \equiv \sum_1^n (Y_i - a - b(x_i - \bar{x}))^2$. Let

Then $\epsilon_i \equiv Y_i - (\hat{\alpha} + \hat{\beta}(x_i - \bar{x}))$ = observation$_i$ − fitted value$_i$ ≡ residual$_i$.

$$SS_E \equiv \sum_1^n r_i^2 \cong \sigma^2 \chi^2_{n-2} \text{ and } \sqrt{SS_{xx}}(\hat{\beta} - \beta) \Big/ \sqrt{\frac{SS_E}{n-2}} \cong \mathcal{T}_{n-2}$$

ESTIMATION

DEF 1. A statistic is any function $T \equiv T(\underline{X})$ of the observations $\underline{X}$. The set Θ of all possible values of the parameter θ in our model $f_\theta(\cdot)$ for the distribution of the observations is called the parameter space. Our interest may be in θ itself, or some $g(\theta)$ [like θ^2, or $\theta_1 + 1.96\,\theta_2$]. Suppose now that $g(\theta)$ is what we wish to estimate. If

(1) $E_\theta(T(\underline{X})) = g(\theta)$ no matter which value of θ is true, then $T(X)$ (or just T, for short) is an unbiased estimator of $g(\theta)$.

The quantity

(2) $Bias(\theta) \equiv E_\theta(T(\underline{X})) - g(\theta)$ is called the bias.

The quantity

(3) $MSE(\theta) \equiv E_\theta\{[T(\underline{X}) - g(\theta)]^2\}$ is called the mean squared error. □

How much does estimator miss the correct value

When T is unbiased for $g(\theta)$, then $MSE(\theta) = Var_\theta[T(\underline{X})]$. In general,

(4) $MSE(\theta) = Var_\theta[T(\underline{X})] + Bias^2(\theta)$,

since $MSE(\theta) = E_\theta\{([T(X) - \mu_T(\theta)] + [\mu_T(\theta) - g(\theta)])^2\}$

$= E_\theta\{[T(X) - \mu_T(\theta)]^2\} + [\mu_T(\theta) - g(\theta)]^2 + 2[\mu_T(\theta) - g(\theta)]\,E[T(\underline{X}) - \mu_T(\theta)]$

$= Var_\theta[T(\underline{X})] + Bias^2(\theta) + 2 \cdot Bias(\theta) \cdot 0$.

EXAMPLE 1 Suppose that $X_1, \ldots, X_n$ are iid (μ, σ^2) (and suppose only this). Then $\bar{X}_n$ is an unbiased estimator of μ with $MSE = Var[\bar{X}_n] = \frac{\sigma^2}{n}$. And $S_n^2 \equiv \frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2$ is unbiased for σ^2, with $MSE = Var[S_n^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$ if we add the assumption that $\mu_4 \equiv E(X-\mu)^4 < \infty$. Are these always good estimators? Can we do better in a more specialized model?

Note that $MSE(\theta) = E_\theta\{[T(\underline{X}) - g(\theta)]^2\}$ is a reasonal criterion to use when comparing two estimators.

Are there any methods/principles that can help us formulate good estimators?

METHOD OF MOMENTS ESTIMATORS

What principles do we invoke to find good estimators?

Method of Moments Principle. This gives the MME = method of moments estimator.

Equate population moment(s) to true moment(s) and solve for the estimator(s).

Example 2 Let $X_1, \ldots, X_n$ be iid Bernoulli(p). So we equate $\bar{X}_n$ to p and solve.

So, the MME of p is $\bar{p}_n \equiv \bar{p}_n(\underline{X}) = \bar{X}_n$.

This estimator $\bar{p}_n$ is unbiased for p with $MSE(p) = Var[\bar{p}_n] = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$.

(The MME of p^2 is $\bar{p}_n^2$.)

Example 3 Let $X_1, \ldots, X_n$ be iid Gamma(α, β) with $\theta = (\alpha, \beta)$ and $f_\theta(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta}$ on $[0, \infty)$.

So we equate $\bar{X}_n$ to $\alpha\beta$ and $\overline{X_n^2} \equiv \frac{1}{n}\sum_1^n X_i^2$ to $\alpha(\alpha+1)\beta^2$,

and note that $\hat{\sigma}^2 \equiv \frac{1}{n}\sum_1^n (X_i - \bar{X})^2 = \overline{X^2} - \bar{X}^2$ is equated to $\alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$.

Solving gives $\bar{\beta} = \hat{\sigma}^2/\bar{X}$ and $\bar{\alpha} = \bar{X}^2/\hat{\sigma}^2$

Note. In a model with k parameters, we use the first k moment equat

Principle (one parameter) set $\mu_\theta = \bar{X}$ and solve for θ

(two parameter) set $\mu_\theta = \bar{X}$

$E_\theta X^2 = \overline{X^2}$ $\left(\sigma_\theta^2 = \overline{X^2} - \bar{X}^2 = \hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_i - \bar{X})^2\right.$

EX 1: Binomial ($p = \theta$) and $\bar{X} = \frac{1}{n}\sum_1^n X_i$ = proportion of successes $= \hat{p} = \hat{\theta}$

$\mu_\theta = \theta$, set $\theta = \bar{X}$, solve for θ

EX 2: Gamma(α, β), mean $= \alpha\beta = \bar{X}$, $f_{\alpha,\beta} = \frac{1}{\beta^\alpha}\frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\frac{x}{\beta}}$ for $x \geq 0$

2nd moment $= \alpha(\alpha+1)\beta^2 = \overline{X^2}$ — $\bar{\beta} = \hat{\sigma}^2/\bar{X}$ MMEs

or variance $= \alpha\beta^2 = \hat{\sigma}^2$ — $\bar{\alpha} = \bar{X}^2/\hat{\sigma}^2$

used to $\Gamma(\alpha)$ fits data

"PLANNED" Syllabus for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Ch1: Probability, conditional probability, independence, and rvs
Mar 31 L2...Sec2.1–2.3$^-$: E(·), mean, variance, and (especially) transformations (no mgf)
Apr 01 Q....HW and some catchup lecturing
L3...Sampling with and without replacement
.....
Apr 05 L4...Sec3.1–3.2(Handout): Bernoulli trials; Geometric, NegBinomialT (and Hypergeometric and Discrete Uniform)
L5...Sec3.3(Handout): Poisson process, Poisson, Exponential, Gamma (and Beta and Uniform)
Q ....HW, and finish Poisson process
L6...Sec2.3: Moment generating function (mgf)
.....
Apr 12 L7...More mgf, and CLT pics
L8...Ch4: Joint, marginal, and conditional distributions
Q....Sec4.1: Conditionals and convolution
L9...Sec4.2: Conditional mean and variance
.....
Apr 19 L10..Sec4.2 Independence and Expectation; Beta densities
L11..Sec3.4 Exponential families; list of variance and covariance formulas
Q....MIDTERM I
L12..Discuss Midtem; the intuitive meaning of covariance; more on the var/cov list
.....
Apr 26 L13..Sec4.5: Correlation inequaliaty; the empirical df; finish the var/cov list
L14..Sec5.1–5.2: Properties of $\bar{X}$ and S^2
Q....
Apr 30 L15..Sec4.5: Bivariate Normal Distribution
.....
May 03 L16..Sec5.3: Derivation of Student-T and Fisher/Snedecor/varianceRatio-F
L17..Handout: Moments via matrices
Q....
L18..Sec7.1–7.2:Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)
.....
May 10 L19..Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)MLE:
Three illustrative MLE examples
L20..More MLE; Sec6.1–6.2: Sufficiency
Q....Work HW on MLE
L21..Completenes, Exponential families, ancillarity
.....
May 17 L22..MIDTERM II
L23..Discuss Midterm, and begin anew on Bayes
Q....Bayes estimators
L24..Bayes examples
.....
May 24 L25..Cramer–Rao inequality
L26..Finish Cramer–Rao
Q....
L27..Hypothesis tests
.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28..Likelihood ratio tests
Q....Work examples of LR-tests
June04 L29..Location and scale models
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

Planned Homework for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Chapter1: #[2,5,11,13,14,25,33,38; do not be formal; should be easy];#17,18,21,23,24
Mar 31 L2...Chapter1: #44,53; Chapter2: #4,8c,17,19,24
Apr 01 Q....
L3...Chapter2: #20; Chapter3: #4,6,12,23,24b; Handout: #1

.....
Apr 05 L4...Handout: #2–5
L5...Handout: #6–7
Q ...
L6...

.....
Apr 12 L7...Chapter2: #9,16, (read 11); Chapter3: #22cd
L8...Chapter2: #33abc; Chapter3: #19,22ab
Q....
L9...Chapter2: # 26ab, 28ab, 30ad; Chapter3: #24c

.....
Apr 19 L10..Chapter4 #13,15; HW #1abc in Conditional Epectation handout about $v^2(y)$
L11..Unoficial: Work some off the HW sheet containing nine
Q....MIDTERM
L12..HW on the back of the midterm

.....
Apr 26 L13..Chapter4 #39, 43, 58
L14..Chapter4 #23(long, re transformations–start early), 44
Q....
Apr 30 L15..Chapter4 # 27, 30, 45
.....
May 03 L16..Chapter4 #19, 20(discuss #21 as well as a certain in class example in this context), 54
L17..Chapter4 #42, 52
Q....Covariance handout #1,2,4,6
L18..MLE handout #B2,B3

.....
May 10 L19..MLE handout #B1,B5
L20..MLE handout #B6,B7,B8
Q....More HW
L21..More HW

.....
May 17 L22..MIDTERM II
L23..
Q....
L24..

.....
May 24 L25..
L26..
Q....
L27..

.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28..
Q
June04 L29..
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

MAXIMUM LIKELIHOOD ESTIMATORS

MLE's (Maximum Likelihood Estimators)

Let us agree to call the joint density function

(5) $L_{\underline{X}}(\theta) \equiv f_\theta(\underline{x})$ $\left(= \prod_{k=1}^{n} f_\theta(x_k) \text{ when } X_1, \ldots, X_n \text{ are iid } f_\theta(\cdot)\right)$

the observed likelihood function of the observed data $\underline{x}$ when we think of the joint density expression as a function of the parameter value θ while holding the experimental outcome $\underline{x}$ fixed. Then

(6) $l_\theta \equiv l(\theta) \equiv l(\theta; \underline{x}) \equiv \log f_\theta(\underline{x})$ $\left(= \sum_{k=1}^{n} \log f_\theta(x_k) = \sum_{k=1}^{n} l(\theta; x_k) \text{ in iid case}\right)$

will be called the observed log likelihood function. Then the maximum likelihood estimate of θ is that value $\hat{\theta}_n \equiv \hat{\theta}(\underline{x})$ that maximizes the observed likelihood for the data value $\underline{x}$ of the experiment $\underline{X}$. (This is highly intuitive, due to R.A. Fisher, and works well.) Meanwhile, $l_\theta \equiv l(\theta; \underline{X})$ is called the theoretical log likelihood, and a formula/algorithm for the value $\hat{\theta}_n \equiv \hat{\theta}(\underline{X})$ that maximizes this expression is said to produce the maximum likelihood estimator (MLE). It is equivalent to think of $\hat{\theta}_n$ as maximizing either $L_{\underline{x}}(\theta)$ or $l(\theta; \underline{x})$; though their maximum values are different, the point of maximization is the same.]

For Discrete Case $P_\theta(\underline{x})$ probability $\underline{x}$ was observed

$L_{\underline{x}}(\theta) = P_\theta(\underline{x})$ which θ is most likely, which θ gives largest probability

(iid case: $\prod_{i=1}^{n} P_\theta(x_i)$)

$l_{\underline{x}}(\theta) = \log_\theta(\underline{x}) = l_\theta = l(\theta) = l(\theta; \underline{x})$ log likelyhood

EX 4: $X_1, \ldots, X_n$ iid Bernoulli (θ)

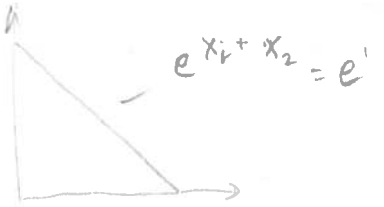
$$L_{\underset{\sim}{X}}(\theta) = P_\theta(\underset{\sim}{x}) = \prod_1^n \theta^{x_i}(1-\theta)^{1-x_i} \cdot \underbrace{1_{\{0,1\}}(x_i)}_{\text{indicator function}}$$

$$= \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i} \qquad \left\{ \prod_1^n 1_{\{0,1\}}(x_i) \right.$$

$$= \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})} \qquad n(\underset{\sim}{x})$$

$$= g_\theta(t) \qquad \text{where } t = \sum_1^n x_i$$

Sufficient statistics is sum of x_i, or $\bar{x}$



Exercise Sheet

Exercises—B

$Q \equiv \sum_1^n i X_i / b_n$ with $b_n \equiv \sum_1^n i^2$.

1. For Example A2, show that Q is the MLE for θ. What is the MLE for θ^3? $X_i \cong N(i\theta, 1)$ and ind.
2. For Example A3, show that R is the MLE for θ. What is the MLE for θ^2? $X_i \cong Exp(mean = i\theta)$ and ind.
3. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim P(\theta)$. What is the MLE of θ? of e^θ? $R \equiv \frac{1}{n}\sum_1^n X_i/i$
4. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim \Gamma(i, \theta)$. What are the MLEs of θ and θ^{-1}?
5. Let $X_1, \ldots, X_n$ be independent, with X_i geometrically distributed with parameter θ. That is, X_i has density $f(x;\theta) = \theta(1-\theta)^x, x = 0, 1, \ldots, 0 < \theta < 1$. Find the MLEs of θ and $(1-\theta)^{-1}$.
6. Let $X_1, \ldots, X_n$ be independently identically distributed with X_i having density $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$. Find the MLE of θ.
7. For Example C in Section 7.1.1, find the MLE of (θ, σ). What is the MLE of $\theta\sigma$? of $\theta\sigma^2$? $X_i \cong N(i\theta, \sigma^2)$;
8. Let (X, Y) have a trinomial distribution with parameters n, θ^2, $2\theta(1-\theta)$, and $(1-\theta)^2$. That is, (X, Y) has joint density function

   $$f(x,y) = k(x,y)(\theta^2)^x(2\theta(1-\theta))^y((1-\theta)^2)^{n-x-y}, \quad x = 0, 1, \ldots, y = 0, 1, \ldots, x + y \leq n.$$

   Suppose n is known. Show that the MLE of θ is $(2X + Y)/2n$.
9. Let (X, Y) have a trinomial distribution with parameters n, θ_1, and θ_2. Show that the MLE of (θ_1, θ_2) is $(X/n, Y/n)$.

Exercises—C

1. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(\theta, \theta^2)$. Find the MLE of θ. (Be sure to verify which root of the quadratic leads to the maximum.)
2. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(\mu, \delta)$. That is, X_i is normally distributed with mean μ and variance δ. Hence, $\theta = (\mu, \delta)$ for this model. Without using the invariance principle, show that the MLE for (μ, δ) is given by $\hat{\mu} = \overline{X}, \hat{\delta} = (n-1)S^2/n$.

13-5-2004

Example 4. Let $X_1, \ldots, X_n$ be iid Bernoulli(θ). The observed likelihood is

(a) $L_{\underline{x}}(\theta) = f_\theta(\underline{x}) = \prod_{k=1}^n \{\theta^{x_k}(1-\theta)^{1-x_k} 1_{\{0,1\}}(x_k)\} = \theta^{\sum_1^n x_k}(1-\theta)^{n-\sum_1^n x_k}\left(\prod_1^n 1_{\{0,1\}}(x_k)\right)$

(b) $= \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}\left\{\prod_{k=1}^n 1_{\{0,1\}}(x_k)\right\}$

(c) $\equiv g_\theta(t) \cdot h(\underline{x})$ where $t \equiv T(\underline{x}) \equiv \sum_{k=1}^n x_k = n\bar{x}_n$

The theoretical log likelihood is will show that $T \equiv \bar{X}_n$ is sufficient for θ

$$\ell(\theta; \underline{X}) = \ell_{\underline{X}}(\theta) = \log L_{\underline{X}}(\theta) = \log g_\theta(T) + \log h(\underline{X})$$

(d) $= n\bar{x}\log\theta + n(1-\bar{x})\log(1-\theta) + \log h(\underline{x})$.

This is maximized by solving

$$0 = \dot{\ell}(\theta; \underline{X}) \equiv \dot{\ell}_{\underline{X}}(\theta) \equiv \frac{\partial}{\partial\theta}\log f_\theta(\underline{X}) = \frac{n\bar{x}}{\theta} - \frac{n(1-\bar{x})}{1-\theta} + 0 = \frac{n\bar{x}(1-\theta) - n\theta(1-\bar{x})}{\theta(1-\theta)}$$

(e) $= \frac{n}{\theta(1-\theta)}[\bar{x} - \theta]$ (for later, $\equiv A_n(\theta)[T(\underline{X}) - T_0(\theta)] = nA(\theta)[T(\underline{x}) - T_0(\theta)]$).

The solution is obviously

(f) $\hat{\theta} \equiv \hat{\theta}_n \equiv \hat{\theta}_n(\underline{X}) = \bar{X} = \bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$.

[Comment: This proof is not valid if $\sum_1^n x_i$ equals 0 or n (i.e., if $\bar{x}_n$ equals 0 or 1) as $L_{\underline{x}}(\theta)$ is this monotone. But it is trivial that the MLE is still $\bar{x}_n$.] Note,

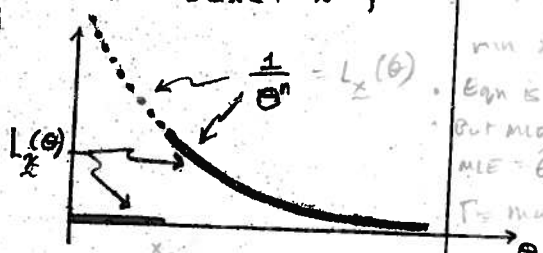
(g) $\sqrt{n}(\bar{X}_n - \theta) \to_d N(0, \theta(1-\theta))$.

Example 5. Let $X_1, \ldots, X_n$ be iid Uniform$(0, \theta)$. The observed likelihood is

(a) $L_{\underline{x}}(\theta) = f_\theta(\underline{x}) = \prod_{k=1}^n \left\{\frac{1}{\theta} 1_{[0,\theta]}(x_k)\right\} = \left\{\frac{1}{\theta^n} 1_{[0,\theta]}\left(\max_{1\le k\le n} x_k\right)\right\} \times \left\{1_{[0,\infty)}\left(\min_{1\le k\le n} x_k\right)\right\}$

(b) $= \{g_\theta(t)\} \times \{h(\underline{x})\}$ with $t \equiv T(\underline{x}) \equiv \left[\max_{1\le k\le n} x_k\right]$;

will show that $T \equiv T(\underline{x}) \equiv \left[\max_{1\le k\le n} x_k\right]$ is sufficient for θ.



From the picture, the MLE of θ is

(c) $\hat{\theta} \equiv \hat{\theta}_n \equiv \hat{\theta}_n(\underline{X}) = \left[\max_{1\le k\le n} X_k\right]$.

(This example is non-standard (or, non-regular) in that we did not solve $0 = \dot{\ell}(\theta; \underline{x})$. It is more specifically non-regular in that the domain of $f_\theta(\underline{x})$ depends on θ.) Now

(d) $\bar{\theta}_n = 2\bar{X}_n$ is the MME of θ, since $\mu_\theta \equiv E_\theta X = \frac{\theta}{2}$. Also,

(e) $\bar{\theta}_n$ has mean θ and variance $\frac{1}{3n}\theta^2$.

How about the MLE $\hat{\theta}_n \equiv T \equiv \left[\max_{1\le k\le n} X_k\right]$? Well,

(f) $F_T(t) = P(T \le t) = P(\max \le t) = P(X \le t)^n = \left(\frac{t}{\theta}\right)^n$; so $f_T(t) = \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1}$ on $[0,\theta]$

Thus

(g) $E_\theta T^a = \int_0^\theta t^a \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1} dt = \frac{n}{n+a}\theta^a$.

Thus

(h) $\frac{n+1}{n}\hat{\theta}_n$ is unbiassed, with variance $\frac{\theta^2}{n(n+2)}$. (This is non-regular; too small)

also

(i) $P(n(\theta - \hat{\theta}_n) > t) = P(\hat{\theta}_n < \theta - \frac{t}{n}) = \left(1 - \frac{t/\theta}{n}\right)^n \to e^{-t/\theta}$. (This is non-regular; not Normal)

Example 6. Let $X_1, \ldots, X_n$ be iid Gamma(α, β). The observed likelihood is

(a) $$L_{\underline{x}}(\theta) = f_\theta(\underline{x}) = \prod_{k=1}^n \left\{ \frac{x_k^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-x_k/\beta} 1_{[0,\infty)}(x_k) \right\} \quad \text{with } \theta \equiv (\alpha, \beta)'$$

$$= \frac{\left(\prod_{k=1}^n x_k\right)^{\alpha-1}}{\beta^{n\alpha}\Gamma^n(\alpha)} e^{-\frac{1}{\beta}\sum_{k=1}^n x_k} \times 1_{[0,\infty)}\left(\min_{1\le k\le n} x_k\right)$$

$$= \frac{e^{(\alpha-1)\sum_{k=1}^n \log x_k}}{\beta^{n\alpha}\Gamma^n(\alpha)} e^{-\frac{1}{\beta}\sum_{k=1}^n x_k} 1_{[0,\infty)}\left(\min_{1\le k\le n} x_k\right)$$

t_1 t_2

(b) $$= \frac{e^{(\alpha-1) n \overline{\log x}}}{\beta^{n\alpha}\Gamma^n(\alpha)} e^{-\frac{1}{\beta} n \bar{x}} \left\{ 1_{[0,\infty)}\left(\min_{1\le k\le n} x_k\right) \right\}$$

(c) $$= g_\theta(t) \times \{h(\underline{x})\} \quad \text{with } t' \equiv (t_1, t_2) = (\overline{\log x}, \bar{x})$$ being sufficient for $\theta' \equiv (\alpha, \beta)$.

The theoretical log likelihood is

(d) $$\ell(\theta; \underline{X}) = -n\alpha \log\beta - n\log\Gamma(\alpha) + (\alpha-1) n \overline{\log X} - \frac{n}{\beta}\bar{X} \ldots$$

We seek to solve the likelihood equations:

(e) $$0 = \dot{\ell}_\beta \equiv \frac{\partial \ell}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{n}{\beta^2}\bar{X} = \frac{n}{\beta^2}\left[\bar{X} - \alpha\beta\right]$$

and

(f) $$0 = \dot{\ell}_\alpha \equiv \frac{\partial \ell}{\partial \alpha} = n\overline{\log X} - n\log\beta - n\frac{d}{d\alpha}\log\Gamma(\alpha) = n\left\{\overline{\log X} - \log\beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right\}.$$

Equation (e) has solution

(g) $$\hat{\beta} \equiv \hat{\beta}_n \equiv \bar{X}_n / \hat{\alpha}_n$$

where $\hat{\alpha} \equiv \hat{\alpha}_n$ is (from equation (f)) the solution of

(h) $$R_n \equiv \log\bar{X} - \overline{\log X} = \log\alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \equiv \log\alpha - \psi(\alpha) \equiv \phi(\alpha).$$

Here, $\psi(\alpha) \equiv \frac{d}{d\alpha}\log\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ is the venerable diagamma function. Also,

(i) $$R_n = \log\bar{X} - \overline{\log X} = \log\bar{X} - \log\left(\prod_1^n X_i\right)^{1/n} = \log\frac{\bar{X}}{\left(\prod_1^n X_i\right)^{1/n}} = \log\frac{\text{Arithmetic Mean}}{\text{Geometric Mean}} > 0.$$

Moreover, it is known that

(j) $\phi(\alpha) \equiv \log\alpha - \psi(\alpha)$ is a strictly ↓ function with range $(0, \infty)$;

thus $\hat{\alpha} \equiv \hat{\alpha}_n$ necessarily exists. (Also, $\phi(\alpha) = \frac{1}{\alpha} \cdot \alpha\phi(\alpha)$, where $\alpha\phi(\alpha)$ is a much more "placid" function than $\phi(\alpha)$, and is easier to interpolate.)

Though we cannot determine a formula for $\hat{\alpha}_n$, it can be shown that

(k) $$\begin{bmatrix} \sqrt{n}(\hat{\alpha}_n - \alpha) \\ \sqrt{n}(\hat{\beta}_n - \beta) \end{bmatrix} \to_d N(0, \Sigma), \quad \text{for an appropriate } \Sigma$$

converges to bivariate normal

□

Chapter Comments

Miscellanea: These appear at the end of most chapters. Rarely will I mention them.
Chapter 2:
Not responsible for sec2.4. (It is good mathematics.)
Chapter 3:
Low priority to Theorem 3.4.3 on page 112.
Sec3.6 may be omitted for now. We will go back as required (not much).
Chapter 4:
Sec4.3: My job is to show you an easier way for much of this.
Sec4.7 may be omitted for now. We will go back as required (not much).
Chapter 5: Lemma 5.3.3 just says that linear combinations of normals are normal, and all you have to do is compute the mean, variance, and covariance. If the covariance of the new guys is 0, then they are independent.
The rest of Sections 5.1-5.3 is straightforward– and is also covered in my handouts.
Section 5.4 on order statistics is important, secondarily. The authors give you mathematical proofs. I briefly showed you the key step to that–but gave you the probabilistic proof instead.
Section 5.5 on limit theorems. We have practiced the CLT much more than this. The rest of this is very secondary–except for the delta method, that we will do later when it is needed.
Section 5.6 on generating rvs is readable. Our main concern is Example 3.6.3, with applications to the Exponential on page 247, chisquare on page 248, and to the normal on page 249. (We stopped there.)
Chapter 6: Sections 6.1 and 6.2 will be interlaced with chapter 7. Sections 6.3 and 6.4 may be omitted. (But, Section 6.3 is *very* well written, and interesting.)
Chapter 7: For Midterm 2, only MME and MLE. Example 7.2.8 is less important. Theorem 7.2.10 is absolutely trivial–with no notation needed. For us, Example 7.2.12 is just a reference. We will do Bayes estimators much as done here, but later. (Sections 7.2.4(EM Algorithm) and 7.3.4 may be omitted. Omitting Satterwaite Example 7.2.3 is ok for us.) We really began with Section 7.3 on Bias and MSE. Omit Example 7.3.6. Crame'r-Rao will be very important later. So will Rao-Blackwell. Omit Section 7.3.4 on loss function optimality, except Example 7.3.9.
Chapter 8 All of Chapter 8 is covered (only lightly on 8.3.3, and omitting 8.3.5)
Chapter 9 We want the ideas of Chapter 9, but it is not very well written. I've never read a well written chapter on this topic.

14-5-2004

Exercise Sheet

Exercises—B

1. For Example A2, ~~show that~~ determine Q is the MLE for θ. What is the MLE for θ^3? $X_i \cong N(i\theta, 1)$ and ind.
2. For Example A3, ~~show that~~ determine R is the MLE for θ. What is the MLE for θ^2? $X_i \cong \text{Exp}(\text{mean} = i\theta)$ and ind
3. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim P(\theta)$. What is the MLE of θ? of e^{θ}?
4. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim \Gamma(i, \theta)$. What are the MLEs of θ and θ^{-1}?
5. Let $X_1, \ldots, X_n$ be independent, with X_i geometrically distributed with parameter θ. That is, X_i has density $f(x;\theta) = \theta(1-\theta)^x, x = 0, 1, \ldots, 0 < \theta < 1$. Find the MLEs of θ and $(1-\theta)^{-1}$.
6. Let $X_1, \ldots, X_n$ be independently identically distributed with X_i having density $f(x;\theta) = \theta x^{\theta - 1}, 0 < x < 1, \theta > 0$. Find the MLE of θ.
7. For Example C in Section 7.1.1, find the MLE of (θ, σ). What is the MLE of $\theta\sigma$? of $\theta\sigma^2$? $X_i \cong N(i\theta, \sigma^2)$
8. Let (X, Y) have a trinomial distribution with parameters n, θ^2, $2\theta(1-\theta)$, and $(1-\theta)^2$. That is, (X, Y) has joint density function

   $$f(x,y) = k(x,y)(\theta^2)^x(2\theta(1-\theta))^y((1-\theta)^2)^{n-x-y}, \quad x = 0, 1, \ldots, y = 0, 1, \ldots, x + y \leq n.$$

   Suppose n is known. Show that the MLE of θ is $(2X + Y)/2n$.
9. Let (X, Y) have a trinomial distribution with parameters n, θ_1, and θ_2. Show that the MLE of (θ_1, θ_2) is $(X/n, Y/n)$.

Exercises—C

1. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(\theta, \theta^2)$. Find the MLE of θ. (Be sure to verify which root of the quadratic leads to the maximum.)
2. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(\mu, \delta)$. That is, X_i is normally distributed with mean μ and variance δ. Hence, $\theta = (\mu, \delta)$ for this model. Without using the invariance principle, show that the MLE for (μ, δ) is given by $\hat{\mu} = \overline{X}, \hat{\delta} = (n-1)S^2/n$.

Ex 1

$X_1, \ldots, X_n$ are iid $N(\mu + \nu, \sigma^2)$

$Y_1, \ldots, Y_{2n}$ are iid $N(\mu - \nu, \sigma^2)$

Care about μ. want to estimate μ / variance of $\bar{x}$

$\hat{\mu} = \frac{\overline{X}_n + \overline{Y}_{2n}}{2}$ has mean μ and variance $\sigma^2 \frac{1}{4}\left(\frac{1}{n} + \frac{1}{2n}\right) = \frac{3}{8}\frac{\sigma^2}{n}$

So $Z \equiv \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\frac{3}{8}}\sigma} \cong N(0,1)$

$SS_{xx} \cong \sigma^2 \chi^2_{n-1}$

$SS_{yy} \cong \sigma^2 \chi^2_{2n-1}$

So $S_p^2 \equiv \frac{SS_{xx} + SS_{yy}}{3n-2} \cong \sigma^2 \frac{\chi^2_{3n-2}}{3n-2}$

Thus $T \equiv \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\frac{3}{8}}S_p} \cong \sigma_{3n-2}$.

CI for μ is $\hat{\mu} \pm t^{(\alpha/2)}_{3n-2} \frac{S_p}{\sqrt{n}}\sqrt{\frac{3}{8}}$

Optimally!

Ex 2

$X_1, \ldots, X_{2n}$ $N(3\mu, 3\sigma^2)$

$Y_1, \ldots, Y_{3n}$ $N(\mu, 5\sigma^2)$

now need to make σ normalized to 1

sd-e for σ

Ex 3 X_i^2 Y_i^2 and Z_i^2's with $\overline{Y}$ and SS_{zz} Post

info lives in $\bar{X}$, $\bar{Y}$, SS_{xx}, SS_{yy}, S_x^2, S_y^2

Review for Midterm II.

Cov calcs and $-1 \le \rho \le 1$.
I give you μ's σ's and ρ. You draw me a scatter cloud of data.

All about $\bar{X}$ and S^2 and student T-technology.
With student T and Snedecor F; via $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ and $\frac{U/m}{V/n}$.

Elementary CI for μ

Bias, MSE, Var
MME's and MLE's

Bivariate Normal
$Y|X=x \sim N(m(x) = \mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x-\mu_X),\ v^2(x) = \sigma^2_{\varepsilon} = \sigma_Y^2(1-\rho^2))$
Regression line } meanings
St Dev line }
Draw the ellipses

$EY = Em(X)$
$\mathrm{Var}[Y] = \mathrm{Var}[m(X)] + E(v^2(X))$.

Linear combs of normals are normal.
Just compute the means and variances
Normals are ind iff covariance is 0.

Minimizing via guess and verify, using cross product is zero.

Example 7. Let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim N(\mu_i(\theta), \sigma^2) = N(\mu_i(\alpha,\beta), \sigma^2)$.

The observed likelihood is

(a) $$L_{\underset{\sim}{Y}}(\theta) = f_\theta(\underset{\sim}{y}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \mu_i(\alpha,\beta))^2} \qquad (\text{with } \theta \equiv (\alpha,\beta)'$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [y_i - \mu_i(\alpha,\beta)]^2}$$

(b) $$\equiv \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} H_{\underset{\sim}{y}}(\alpha,\beta)}$$

with $H_{\underset{\sim}{y}}(\alpha,\beta) \equiv \sum_{i=1}^n [y_i - \mu_i(\alpha,\beta)]^2$

The theoretical log likelihood is thus

(c) $$\ell_\theta = \ell_\theta(\underset{\sim}{Y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} H_{\underset{\sim}{Y}}(\alpha,\beta) .$$

Thus the likelihood equation for σ^2 is

(d) $$0 = \dot{\ell}_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} H_{\underset{\sim}{Y}}(\alpha,\beta) = \frac{n}{2\sigma^4}\left[\frac{1}{n} H_{\underset{\sim}{Y}}(\alpha,\beta) - \sigma^2\right].$$

Regardless of the value of $\theta = (\alpha,\beta)'$, this $0 = \dot{\ell}_{\sigma^2}$ equation has solution $\frac{1}{n} H_{\underset{\sim}{Y}}(\alpha,\beta)$; so it follows the MLE for σ^2 is

$$\hat{\sigma}^2 \equiv \frac{1}{n} H_{\underset{\sim}{Y}}(\hat{\theta})$$

for the MLE $\hat{\theta}$ of θ.

Note that $L_{\underset{\sim}{Y}}(\theta)$ in (b) is maximized by that value $\hat{\theta}$ of θ that minimizes the sum of squares $H_{\underset{\sim}{Y}}(\theta) \equiv \sum_{i=1}^n [Y_i - \mu_i(\theta)]^2$. That is, the MLE of θ is also the least squares estimator (the LSE) of θ.

<u>One Sample Problem</u>.

$$H_{\underset{\sim}{Y}}(\theta) = \sum_{i=1}^n [Y_i - \mu]^2, \text{ so } \hat{\mu} = \bar{Y}$$

<u>Two Sample Problem</u>

$$H_{\underset{\sim}{Y}}(\theta) = \sum_{i=1}^n (X_i - \mu)^2 + \sum_{j=1}^n (Y_j - \nu)^2, \text{ so } \hat{\mu} = \bar{X} \text{ and } \hat{\nu} = \bar{Y}.$$

<u>Regression Model</u>

$$H_{\underset{\sim}{Y}}(\theta) = \sum_{i=1}^n [Y_i - (\alpha + \beta(x_i - \bar{x}))]^2, \text{ and } \hat{\alpha} = \bar{Y} \text{ with } \hat{\beta} = \frac{\sum_1^n (x_i - \bar{x}) Y_i}{\sum_1^n (x_i - \bar{x})^2}$$

will be shown.

Shorack STAT 506 May 17, 2004

Second Midterm (closed book and closed notes).

#1. Let $Y_1, \ldots, Y_n$ be independent having the Rayleigh(θ) density

$$f_\theta(y) \equiv \frac{y}{\theta^2} \exp(-\frac{y^2}{2\theta^2}) \qquad \text{for } y \geq 0.$$

(a) (25) Determine the MLE $\hat{\theta}_n$ of θ.
(b) (8) In a different vein, determine the median of the Rayleigh(θ) distribution. Hint let $Z = \frac{Y^2}{2\theta^2}$ appeal to gamma

#2. (33) Let $X_1, \ldots, X_{2n}$ be iid $N(\mu - 3\nu, 2\sigma^2)$. Let $Y_1, \ldots, Y_{3n}$ be iid $N(\mu + \nu, 3\sigma^2)$.
(a) Determine some unbiased estimator $\tilde{\mu}$ of μ (that uses both $\bar{X}_{2n}$ and $\bar{Y}_{3n}$ in its definition).
(b) Find the exact distribution of $\tilde{\mu}$.
(c) Determine an appropriate estimator of σ^2 (base it on data from both samples).
(d) Define an appropriate Student-$\mathcal{T}$ statistic based on $\tilde{\mu}$, and use it to find a $(1-\alpha)$-confidence interval for μ.

#3. (a) (25) In the context of problem #1, find the MME $\bar{\theta}_n$ of θ.
(b) (9) Determine the asymptotic distribution of $\sqrt{n}\,(\bar{\theta}_n - \theta)$.

THE EXAM IS OVER. WHAT FOLLOWS IS HOMEWORK

#2. (e) Determine the MLEs $\hat{\mu}$, $\hat{\nu}$, and $\hat{\sigma}^2$.
(f) Find the efficiency defined via $\lim \mathrm{Var}[\hat{\mu}]/\mathrm{Var}[\tilde{\mu}]$.
Replace the basic assumptions by the following.
Let $X_1, \ldots, X_{2n}$ be iid $N(\mu - 3\nu, \sigma^2)$. Let $Y_1, \ldots, Y_{3n}$ be iid $N(\mu + \nu, \tau^2)$.
(g) Find an estimator of $\Delta^2 \equiv \tau^2/\sigma^2$ that has an $\mathcal{F}$–distribution. test about ratio of variances
(h) Use this to give a $1 - \alpha$ confidence interval for $\Delta \equiv \tau/\sigma$.

#3. (c) (b) Determine the asymptotic distribution of $\sqrt{n}\,(\tilde{\theta}_n^2 - \theta^2)$.
(d) Determine the exact distribution of $Y^2/(2\theta^2)$.
(e) Determine the exact distribution of $\hat{\theta}_n^2$.
(f) Find the efficiency defined via $\lim \mathrm{Var}[\hat{\mu}]/\mathrm{Var}[\bar{\mu}]$.

"PLANNED" Syllabus for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Ch1: Probability, conditional probability, independence, and rvs
Mar 31 L2...Sec2.1–2.3^{-}: E(·), mean, variance, and (especially) transformations (no mgf)
Apr 01 Q....HW and some catchup lecturing
L3...Sampling with and without replacement
.....
Apr 05 L4...Sec3.1–3.2(Handout): Bernoulli trials; Geometric, NegBinomialT (and Hypergeometric and Discrete Uniform)
L5...Sec3.3(Handout): Poisson process, Poisson, Exponential, Gamma (and Beta and Uniform)
Q ....HW, and finish Poisson process
L6...Sec2.3: Moment generating function (mgf)
.....
Apr 12 L7...More mgf, and CLT pics
L8...Ch4: Joint, marginal, and conditional distributions
Q....Sec4.1: Conditionals and convolution
L9...Sec4.2: Conditional mean and variance
.....
Apr 19 L10..Sec4.2 Independence and Expectation; Beta densities
L11..Sec3.4 Exponential families; list of variance and covariance formulas
Q....MIDTERM I
L12..Discuss Midtem; the intuitive meaning of covariance; more on the var/cov list
.....
Apr 26 L13..Sec4.5: Correlation inequaliaty; the empirical df; finish the var/cov list
L14..Sec5.1–5.2: Properties of $\bar{X}$ and S^2
Q....
Apr 30 L15..Sec4.5: Bivariate Normal Distribution
.....
May 03 L16..Sec5.3: Derivation of Student-T and Fisher/Snedecor/varianceRatio-F
L17..Handout: Moments via matrices
Q....
L18..Sec7.1–7.2:Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)
.....
May 10 L19..Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)MLE:
Three illustrative MLE examples (Did Bernoulli(θ) and Uniform$(0,\theta)$).
L20..More MLE.
Q....Student-T technology.
L21..MLEs for Gamma(α,β) and various Normal$(\mu(\theta),\sigma^2)$ examples
.....
May 17 L22..MIDTERM II
L23..Discuss Midterm
Q....Bayes estimators
L24..Bayes examples
.....
May 24 L25..Cramer–Rao inequality
L26..Finish Cramer–Rao
Q....
L27..Hypothesis tests
.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28..Likelihood ratio tests
Q....Work examples of LR-tests
June04 L29..Location and scale models
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

Planned Homework for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Chapter1: #[2,5,11,13,14,25,33,38; do not be formal; should be easy];#17,18,21,23,24
Mar 31 L2...Chapter1: #44,53; Chapter2: #4,8c,17,19,24
Apr 01 Q....
L3...Chapter2: #20; Chapter3: #4,6,12,23,24b; Handout: #1
.....
Apr 05 L4...Handout: #2–5
L5...Handout: #6–7
Q ...
L6...
.....
Apr 12 L7...Chapter2: #9,16, (read 11); Chapter3: #22cd
L8...Chapter2: #33abc; Chapter3: #19,22ab
Q....
L9...Chapter2: # 26ab, 28ab, 30ad; Chapter3: #24c
.....
Apr 19 L10..Chapter4 #13,15; HW #1abc in Conditional Epectation handout about $v^2(y)$
L11..Unoficial: Work some off the HW sheet containing nine
Q....MIDTERM
L12..HW on the back of the midterm
.....
Apr 26 L13..Chapter4 #39, 43, 58
L14..Chapter4 #23(long, re transformations–start early), 44
Q....
Apr 30 L15..Chapter4 # 27, 30, 45
.....
May 03 L16..Chapter4 #19, 20(discuss #21 as well as a certain in class example in this context), 54
L17..Chapter4 #42, 52
Q....Covariance handout #1,2,4,6
L18..MLE handout #B2,B3
.....
May 10 L19..MLE handout #B1,B5
L20..MLE handout #B6,B7,B8
Q....
L21..
.....
May 17 L22..MIDTERM II
L23..Exercises on the back of Midterm II
Q....Chapter7 #24,23;
Find the Bayes estimators of $\theta(1-\theta)$ and $\theta/(1-\theta)$ in the Bernoulli(θ) Example 1.
L24.. Ch 7 22
.....
May 24 L25..
L26..
Q....
L27..
.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28..
Q
June04 L29..
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

Shorack STAT 506 May 17, 2004

Second Midterm (closed book and closed notes).

#1. Let $Y_1, \ldots, Y_n$ be independent having the Rayleigh(θ) density

$$f_\theta(y) \equiv \frac{y}{\theta^2} \exp(-\frac{y^2}{2\theta^2}) \qquad \text{for } y \geq 0.$$

(a) (25) Determine the MLE $\hat{\theta}_n$ of θ.
(b) (8) In a different vein, determine the median of the Rayleigh(θ) distribution.
Hint: Substitute $z = y^2/(2\theta^2)$, both here and when needed below.

#2. (33) Let $X_1, \ldots, X_{2n}$ be iid $N(\mu - 3\nu, 2\sigma^2)$. Let $Y_1, \ldots, Y_{3n}$ be iid $N(\mu + \nu, 3\sigma^2)$.
(a) Determine some unbiased estimator $\tilde{\mu}$ of μ (that uses both $\bar{X}_{2n}$ and $\bar{Y}_{3n}$ in its definition).
(b) Find the exact distribution of $\tilde{\mu}$.
(c) Determine an appropriate estimator of σ^2 (base it on data from both samples).
(d) Define an appropriate Student-$\mathcal{T}$ statistic based on $\tilde{\mu}$, and use it to find a $(1-\alpha)$-confidence interval for μ.

#3. (a) (25) In the context of problem #1, find the MME $\bar{\theta}_n$ of θ.
(b) (9) Determine the asymptotic distribution of $\sqrt{n}\,(\bar{\theta}_n - \theta)$.

THE EXAM IS OVER. WHAT FOLLOWS IS HOMEWORK

#2. (e) Determine the MLEs $\hat{\mu}$, $\hat{\nu}$, and $\hat{\sigma}^2$.
(f) Specify an unbiased estimator of σ^2 that is similar to the MLE $\hat{\sigma}^2$.

Replace the basic assumptions by the following.
Let $X_1, \ldots, X_{2n}$ be iid $N(\mu - 3\nu, \sigma^2)$. Let $Y_1, \ldots, Y_{3n}$ be iid $N(\mu + \nu, \tau^2)$.
(g) Find an estimator of $\Delta^2 \equiv \tau^2/\sigma^2$ that has an $\mathcal{F}$–distribution.
(h) Use this to give a $1 - \alpha$ confidence interval for $\Delta \equiv \tau/\sigma$.

#3. (d) Determine the exact distribution of $Y^2/(2\theta^2)$.
(e) Determine the exact distribution of $\hat{\theta}_n^2$.
(f) Find the efficiency defined via $\lim_n \mathrm{Var}[\hat{\theta}]/\mathrm{Var}[\bar{\theta}]$.

#4. Let $X_1, \ldots, X_{2n}$ be iid $N(\mu, 2\sigma^2)$. Let $Y_1, \ldots, Y_{3n}$ be iid $N(2\mu, 3\sigma^2)$.
(a) Each $\hat{\mu}_c \equiv c\bar{X}_{2n} + (1-c)\frac{1}{2}\bar{Y}_{3n}$ (with $0 \le c \le 1$) is an unbiased estimator of μ. For which value of c does this estimator have minimum variance?
(b) Determine the MLE of μ. It will be the same estimator.

Efficiency: $\frac{n}{m} = \frac{\sigma_1^2}{\sigma_2^2}$ where $\mathrm{Var}[\hat{\theta}_n] = \frac{\sigma_1^2}{n}$

$\mathrm{Var}[\tilde{\theta}_m] = \frac{\sigma_2^2}{m}$

BAYES ESTIMATORS

Classically: Parameter Θ is fixed and unknown. We use $\hat{\Theta}_n(\underline{X})$ to estimate Θ. Also, perhaps a certain random set contains Θ with probability close to 1.

Bayesian: The distribution $\pi(\cdot)$ of the rv Θ represents our prior knowledge of the value Θ in a model $f(\underline{x}|\theta)$. We will base our actions on the posterior distribution $\pi(\cdot|\underline{x})$.

Model The basic experiment X has distribution $f_\Theta(x) = f(x|\theta)$, where the value Θ of the rv Θ is distributed according to the prior distribution $\pi(\cdot)$. So we obtain information on "the actual Θ" via $f_\Theta(\underline{x}) = f(\underline{x}|\theta)$. Then

(1) $$\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\pi(\theta)}{\int f(\underline{x}|\theta')\pi(\theta')d\theta'} \equiv \frac{f(\underline{x}|\theta)\pi(\theta)}{f_{\underline{X}}(\underline{x})} \equiv (\text{the posterior distribution})$$

updates our prior information in light of the observed data $\underline{x}$ (and, according to the model), where

(2) $$f_{\underline{X}}(\underline{x}) \equiv \int f(\underline{x}|\theta')\pi(\theta')d\theta' = (\text{the marginal distribution of } \underline{x}).$$

We now agree that:

$E(\theta|\underline{x}) \equiv$ (the mean of the posterior distribution $\pi(\theta|\underline{x})$)

$\text{Var}[\theta|\underline{x}] \equiv$ (the variance of the posterior distribution $\pi(\theta|\underline{x})$).

Moreover:

(3) $E(\theta|\underline{x})$ would seem to be a reasonable estimator of Θ.

Quantiles of the posterior $\pi(\theta|\underline{x})$ should give a reasonable Bayesian interval for Θ.

$\text{Var}[\theta|\underline{x}]$ would seem to be useful in assessing the estimator $E(\theta|\underline{x})$.

If we now observe more data $\underline{y}$, we just repeat the above scheme:

(4) Let $\pi(\theta|\underline{x})$ be the new prior, and the new posterior will turn out to be $\pi(\theta|\underline{x},\underline{y})$.

"DEF" The **conjugate prior** is the prior that via mathematically nice calculations leads to a posterior in the same family as the prior. [It need not exist; but it does for exponential families. Thus, textbook and lecture examples are of this exponential family type.]

EXAMPLE 1. Suppose we observe $X_1, \ldots, X_n$ independent Bernoulli(θ) rvs. As a *prior* for the distribution $\pi(\cdot)$ on values θ in $(0,1)$, we use Beta(α, β). We will replace $\underline{X}$ by the sufficient statistic $Y \equiv Y(\underline{X}) \equiv \sum_{i=1}^n X_i$. Then

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f_X(y)} = \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \text{Constant}(\alpha,\beta,y)\ \theta^{y+\alpha-1}(1-\theta)^{n+\beta-y-1} \quad \text{for } 0 \le \theta \le 1.$$

$$(5) \qquad = \frac{1}{B(y+\alpha, n+\beta-y)}\theta^{y+\alpha-1}(1-\theta)^{n+\beta-y-1} \quad \text{for } 0 \le \theta \le 1.$$

(must)

This *posterior distribution* $\pi(\theta|y)$ has mean (or *Bayes estimator*)

$$(6) \quad E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n} = \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta} + \left(\frac{n}{\alpha+\beta+n}\right)\frac{y}{n} \quad (\equiv \hat{p}_{Bayes}).$$

Fact: The Beta(α,β) distribution has

$$(7) \quad \text{mean} = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \text{variance} = \frac{1}{\alpha+\beta+1}\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta}.$$

[Thus $\frac{\alpha}{\alpha+\beta}$ is chosen to reflect our prior mean, and $\alpha+\beta$ is chosen to reflect the strength of our apriori belief.]

Note from (6) that the *Bayes estimator* $\hat{p}_{Bayes} \equiv E(\theta|y)$ satisfies

(8) $\hat{p}_{Bayes}$ is a weighted linear combination of the prior mean $\frac{\alpha}{\alpha+\beta}$ and the data mean $\frac{y}{n} = \bar{x}$.

[*Note:* The *conjugate priors* for a Binomial(θ) model is the Beta(α,β) family.]

Remark: The marginal of y can be calculated as

$$f_Y(y) = \binom{n}{y}\frac{\text{Beta}(\alpha+y, n+\beta-y)}{\text{Beta}(\alpha,\beta)} \quad \text{for } 0 \le y \le n,$$

and is called the *beta-binomial distribution*. □

BAYES ESTIMATORS

Classically: Parameter Θ is fixed and unknown. We use $\hat{\Theta}_n(\underline{X})$ to estimate Θ. Also, perhaps a certain random set contains Θ with probability close to 1.

Bayesian: The distribution $\pi(\cdot)$ of the rv Θ represents our prior knowledge of the value Θ in a model $f(\underline{x}|\theta)$. We will base our actions on the posterior distribution $\pi(\cdot|\underline{x})$.

Model The basic experiment X has distribution $f_\Theta(x) = f(x|\theta)$, where the value θ of the rv Θ is distributed according to the prior distribution $\pi(\cdot)$. So we obtain information on "the actual θ" via $f_\theta(\underline{x}) = f(\underline{x}|\theta)$. Then

(1) $$\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\pi(\theta)}{\int f(\underline{x}|\theta')\pi(\theta')d\theta'} \equiv \frac{f(\underline{x}|\theta)\pi(\theta)}{f_{\underline{X}}(\underline{x})} \equiv (\text{the posterior distribution})$$

updates our prior information in light of the observed data $\underline{x}$ (and, according to the model), where

(2) $$f_{\underline{X}}(\underline{x}) \equiv \int f(\underline{x}|\theta')\pi(\theta')d\theta' = (\text{the marginal distribution of } \underline{x}).$$

We now agree that:

$E(\theta|\underline{x}) \equiv$ (the mean of the posterior distribution $\pi(\theta|\underline{x})$

$Var[\theta|\underline{x}] \equiv$ (the variance of the posterior distribution $\pi(\theta|\underline{x})$.

Moreover:

(3) $E(\theta|\underline{x})$ would seem to be a reasonable estimator of Θ.

Quantiles of the posterior $\pi(\theta|\underline{x})$ should give a reasonable Bayesian interval for Θ

$Var[\theta|\underline{x}]$ would seem to be useful in assessing the estimator $E(\theta|\underline{x})$.

If we now observe more data $\underline{y}$, we just repeat the above scheme:

(4) Let $\pi(\theta|\underline{x})$ be the new prior, and the new posterior will turn out to be $\pi(\theta|\underline{x},\underline{y})$.

"DEF" The conjugate prior is the prior that via mathematically nice calculations leads to a posterior in the same family as the prior. [It need not exist; but it does for exponential families. Thus, textbook and lecture examples are of this exponential family type.]

EXAMPLE 1. Suppose we observe $X_1, \ldots, X_n$ independent Bernoulli(θ) rvs. As a *prior* for the distribution $\pi(\cdot)$ on values θ in $(0,1)$, we use Beta(α,β). We will replace $\underset{\sim}{X}$ by the sufficient statistic $Y \equiv Y(\underset{\sim}{X}) \equiv \sum_{i=1}^n X_i$. Then

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f_X(y)} = \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \text{Constant}(\alpha,\beta,y)\ \theta^{y+\alpha-1}(1-\theta)^{n+\beta-y-1} \quad \text{for } 0\le\theta\le 1.$$

$$\overset{\text{must}}{=} \frac{1}{B(y+\alpha, n+\beta-y)}\theta^{y+\alpha-1}(1-\theta)^{n+\beta-y-1} \quad \text{for } 0\le\theta\le 1. \tag{5}$$

This *posterior distribution* $\pi(\theta|y)$ has mean (or *Bayes estimator*)

$$E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n} = \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta} + \left(\frac{n}{\alpha+\beta+n}\right)\frac{y}{n} \quad (\equiv \hat{p}_{\text{Bayes}}). \tag{6}$$

Fact: The Beta(α,β) distribution has

$$\text{mean} = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \text{variance} = \frac{1}{\alpha+\beta+1}\cdot\frac{\alpha}{\alpha+\beta}\cdot\frac{\beta}{\alpha+\beta}. \tag{7}$$

[Thus $\frac{\alpha}{\alpha+\beta}$ is chosen to reflect our prior mean, and $\alpha+\beta$ is chosen to reflect the strength of our apriori belief.]

Note from (6) that the *Bayes estimator* $\hat{p}_{\text{Bayes}} \equiv E(\theta|y)$ satisfies

(8) $\hat{p}_{\text{Bayes}}$ is a weighted linear combination of the prior mean $\frac{\alpha}{\alpha+\beta}$ and the data mean $\frac{y}{n} = \bar{x}$.

[*Note:* The *conjugate priors* for a Binomial(θ) model is the Beta(α,β) family.]

Remark: The marginal of y can be calculated as

$$f_Y(y) = \binom{n}{y}\frac{\text{Beta}(\alpha+y, n+\beta-y)}{\text{Beta}(\alpha,\beta)} \quad \text{for } 0\le y\le n,$$

and is called the *beta-binomial distribution*. □

BAYES ESTIMATORS

Classically: Parameter Θ is fixed and unknown. We use $\hat{\Theta}_n(\underset{\sim}{X})$ to estimate Θ. Also, perhaps a certain random set contains Θ with probability close to 1.

Bayesian: The distribution $\pi(\cdot)$ of the rv Θ represents our prior knowledge of the value Θ in a model $f(\underset{\sim}{x}|\theta)$. We will base our actions on the posterior distribution $\pi(\cdot|\underset{\sim}{x})$.

Model The basic experiment X has distribution $f_\Theta(x) = f(x|\theta)$, where the value Θ of the rv Θ is distributed according to the prior distribution $\pi(\cdot)$. So we obtain information on "the actual Θ" via $f_\Theta(\underset{\sim}{x}) = f(\underset{\sim}{x}|\theta)$. Then

(1) $$\pi(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{\int f(\underset{\sim}{x}|\theta')\pi(\theta')d\theta'} \equiv \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{f_{\underset{\sim}{X}}(\underset{\sim}{x})} \equiv$$ (the posterior distribution updates our prior information in light of the observed data $\underset{\sim}{x}$ (and, according to the model), where

(2) $$f_{\underset{\sim}{X}}(\underset{\sim}{x}) \equiv \int f(\underset{\sim}{x}|\theta')\pi(\theta')d\theta' =$$ (the marginal distribution of $\underset{\sim}{x}$).

We now agree that:

$E(\Theta|\underset{\sim}{x}) \equiv$ (the mean of the posterior distribution $\pi(\theta|\underset{\sim}{x})$

$Var[\Theta|\underset{\sim}{x}] \equiv$ (the variance of the posterior distribution $\pi(\theta|\underset{\sim}{x})$. How spread the posterior dist is

Moreover:

(3) $E(\Theta|\underset{\sim}{x})$ would seem to be a reasonable estimator of Θ.

Quantiles of the posterior $\pi(\theta|\underset{\sim}{x})$ should give a reasonable Bayesian interval for Θ.

$Var[\Theta|\underset{\sim}{x}]$ would seem to be useful in assessing the estimator $E(\Theta|\underset{\sim}{x})$.

If we now observe more data $\underset{\sim}{y}$, we just repeat the above scheme:

(4) Let $\pi(\theta|\underset{\sim}{x})$ be the new prior, and the new posterior will turn out to be $\pi(\theta|\underset{\sim}{x},\underset{\sim}{y})$.

"DEF" The conjugate prior is the prior that via mathematically nice calculations leads to a posterior in the same family as the prior. [It need not exist; but it does for exponential families. Thus, textbook and lecture examples are of this exponential family type.]

Choice of prior is subjective. Try different priors can be used to test robustness of a model.

EXAMPLE 1. Suppose we observe $X_1, \ldots, X_n$ independent Bernoulli(θ) rvs. As a *prior* for the distribution $\pi(\cdot)$ on values θ in $(0,1)$, we use Beta(α, β). We will replace $\underset{\sim}{X}$ by the sufficient statistic $Y \equiv Y(\underset{\sim}{X}) \equiv \sum_{i=1}^n X_i$. Then

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f_X(y)} = \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \text{Constant}(\alpha,\beta,y)\ \theta^{y+\alpha-1}(1-\theta)^{n+\beta-y-1} \quad \text{for } 0 \le \theta \le 1.$$

$$\overset{\text{must}}{=} \frac{1}{B(y+\alpha, n+\beta-y)}\theta^{y+\alpha-1}(1-\theta)^{n+\beta-y-1} \quad \text{for } 0 \le \theta \le 1. \tag{5}$$

This *posterior distribution* $\pi(\theta|y)$ has mean (or *Bayes estimator*)

$$E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n} = \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta} + \left(\frac{n}{\alpha+\beta+n}\right)\frac{y}{n} \quad (\equiv \hat{p}_{\text{Bayes}}). \tag{6}$$

Fact: The Beta(α,β) distribution has

$$\text{mean} = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \text{variance} = \frac{1}{\alpha+\beta+1}\cdot\frac{\alpha}{\alpha+\beta}\cdot\frac{\beta}{\alpha+\beta}. \tag{7}$$

[Thus $\frac{\alpha}{\alpha+\beta}$ is chosen to reflect our prior mean, and $\alpha+\beta$ is chosen to reflect the strength of our apriori belief.]

Note from (6) that the *Bayes estimator* $\hat{p}_{\text{Bayes}} \equiv E(\theta|y)$ satisfies

(8) $\hat{p}_{\text{Bayes}}$ is a weighted linear combination of the prior mean $\frac{\alpha}{\alpha+\beta}$ and the data mean $\frac{y}{n} = \bar{x}$.

[*Note:* The *conjugate priors* for a Binomial(θ) model is the Beta(α,β) family.]

Bayesians would do

$$\pi(\theta|x) \propto \theta^y(1-\theta)^{n-y} \cdot C_{\alpha,\beta}\,\theta^{\alpha-1}(1-\theta)$$

Remark: The marginal of y can be calculated as

$$f_Y(y) = \binom{n}{y}\frac{\text{Beta}(\alpha+y, n+\beta-y)}{\text{Beta}(\alpha,\beta)} \quad \text{for } 0 \le y \le n,$$

and is called the *beta-binomial distribution*. □

CRAMÉR - RAO INEQUALITY

GRS

Notation Let $f_\theta(\cdot)$ denote the "regular" density of our experiment $\underset{\sim}{X} \equiv (X_1, \ldots, X_n)'$. Then:

(1) $\dot{l}_n(\theta; \underset{\sim}{X}) = \frac{\partial}{\partial\theta} \log f_\theta(\underset{\sim}{X})$ $\left(= \sum_{i=1}^n \dot{l}(\theta; X_i) = \sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(X_i) \quad \text{if iid}\right)$

$= (\text{the } \underline{\text{score function}})$ (the score function, for one observation)

(2) $\mathcal{I}_n(\theta) \equiv E_\theta\{[\dot{l}_n(\theta; \underset{\sim}{X})]^2\}$ $\left(= n\mathcal{I}(\theta) = n\, E_\theta\{[\dot{l}(\theta; X)]^2\} \quad \text{if iid}\right)$

$= (\underline{\text{Fisher information}})$ (Fisher information for one observation)

(3) $\bar{\mathcal{I}}_n(\theta) \equiv E_\theta\{-\ddot{l}_n(\theta; \underset{\sim}{X})\}$ $\left(= n\bar{\mathcal{I}}(\theta) = n E_\theta\{-\ddot{l}(\theta; X)\} \quad \text{if iid}\right)$.

Let $T \equiv T(\underset{\sim}{X})$ be a statistic (having $E_\theta|T(\underset{\sim}{X})| < \infty$ for all θ), and suppose both:

(4) $\tau(\theta) \equiv E_\theta\{T(\underset{\sim}{X})\}$ has a derivative on the (open) parameter set Θ.

(5) $E_\theta\{T(\underset{\sim}{X})\}$ and $E_\theta\{1\}$ can both be differentiated under the integral sign.

Theorem 1. (Cramér - Rao Information Inequality) Suppose $f_\theta(\cdot)$ is "regular"; that is, (4) and (5). Then for all θ:

(6) $E_\theta\{\dot{l}_n(\theta; \underset{\sim}{X})\} = 0$.

(7) $\mathrm{Var}_\theta[T(\underset{\sim}{X})] \geq \frac{[\dot{\tau}(\theta)]^2}{\mathcal{I}_n(\theta)}$ $\left(= \frac{[\dot{\tau}(\theta)]^2}{n\,\mathcal{I}(\theta)} \quad \text{if iid}\right)$.

(8) $\mathcal{I}_n(\theta) = \bar{\mathcal{I}}_n(\theta)$ when one can take two derivatives under the integral sign of $E_\theta\{1\}$.

Corollary Equality is achieved in (6) when

(9) $\dot{l}_n(\theta; \underset{\sim}{X}) = \frac{\partial}{\partial\theta} \log f_\theta(\underset{\sim}{X}) = A_n(\theta)[T_0(\underset{\sim}{X}) - \tau_0(\theta)]$

for some constant $A_n(\theta)$, and for some function $\tau_0(\theta)$. (i.e., (9) defines $\tau_0(\theta)$ and $A_n(\theta)$).

In this case, the CRBound (for this particular $T_0(\underset{\sim}{X})$) becomes

(10) $\mathrm{Var}_\theta[T_0(\underset{\sim}{X})] = \left|\frac{\dot{\tau}_0(\theta)}{A_n(\theta)}\right|$ $\left(= \left|\frac{\dot{\tau}_0(\theta)}{n A(\theta)}\right| \quad \text{if iid}\right)$

and $\frac{[\dot{\tau}(\theta)]^2}{\mathcal{I}_n(\theta)} = \frac{\mathcal{I}_n(\theta)}{A_n^2(\theta)}$

(11) $\mathcal{I}_n(\theta) = |\dot{\tau}_0(\theta) A_n(\theta)|$.

Moreover, only for linear functions of $\tau_0(\theta)$ is the CRBound achieved. (For other $T \equiv T(\underset{\sim}{X})$ and other $\tau(\theta)$, the appropriate result is (7).)

Example 1. (Poisson(θ)) Let $X_1, \ldots, X_n$ be iid Poisson(θ). Then

(a) $p_\theta(x) = e^{-\theta}\theta^x/x!$ and $l(\theta; x) = -\theta + x\log\theta - \log x!$

Thus

$$\dot{l}(\theta; x) = -1 + \frac{x}{\theta} = \frac{x-\theta}{\theta} \quad \text{and} \quad -\ddot{l}(\theta; x) = \frac{x}{\theta^2}$$

(b) $\hat{=} (0, \frac{1}{\theta})$ $\hat{=} (\frac{1}{\theta}, \boxtimes)$

Note that $E_\theta[\dot{l}(\theta; X)] = E_\theta(\frac{X-\theta}{\theta}) = 0$, and

$$Q(\theta) = \mathrm{Var}_\theta[\dot{l}(\theta; X)] = \mathrm{Var}_\theta[\tfrac{X-\theta}{\theta}] \quad \text{and} \quad \bar{Q}(\theta) = E_\theta\{-\ddot{l}(\theta; X)\} = E_\theta\{\tfrac{X}{\theta^2}\}$$

(c) $= \frac{\theta}{\theta^2} = \frac{1}{\theta}$ $= \frac{\theta}{\theta^2} = \frac{1}{\theta}$.

Thus the CR Inequality gives: Any unbiased estimator $T \equiv T(X)$ of $\tau(\theta)$ satisfies

$$\mathrm{Var}_\theta[T] \geq \frac{[\tau'(\theta)]^2}{n\,Q(\theta)} = \frac{[\tau'(\theta)]^2}{n\,\bar{Q}(\theta)} = \frac{[\tau'(\theta)]^2}{n\,Q(\theta)} \quad \left(= \frac{[1]^2}{n\cdot\frac{1}{\theta}} = \frac{\theta}{n} \text{ for } \tau_0(\theta) \equiv \theta\right).$$

Now, let us consider $X_1, \ldots, X_n$ iid Poisson(θ). Then:

$$l(\theta, \underset{\sim}{X}) = \log\left(e^{-n\theta}\theta^{\sum_1^n X_i}/\textstyle\prod_1^n X_i!\right) = -n\theta + n\bar{X}\log\theta - \log\prod_1^n X_i!\,.$$

(d) $\dot{l}(\theta, \underset{\sim}{X}) = -n + \frac{n\bar{X}}{\theta} = \frac{n}{\theta}(\bar{X} - \theta) = n\,A(\theta)\left[T_0(\underset{\sim}{X}) - \tau_0(\theta)\right]$

The CRBound for the parameter $\tau_0(\theta) = \theta$ is thus achieved by $T_0(\underset{\sim}{X}) \equiv \bar{X}$, and the variance of $\bar{X}$ is $|\dot{\tau}(\theta)/A_n(\theta)| = \frac{\theta}{n}$. Only for linear functions of $\tau_0(\theta) = \theta$ is it possible to exactly achieve this bound (by the same linear funtion of $T_0(\underset{\sim}{X}) = \bar{X}$).

Note. Solving the likelihood equation

(e) $0 = \dot{l}(\theta; \underset{\sim}{X}) = \frac{n}{\theta}(\bar{X} - \theta)$

shows that $\hat{\theta}_n \equiv \hat{\theta}_n(\underset{\sim}{X}) = \bar{X}$ is the MLE of θ.

Example 2. (Uniform$(0,\theta)$) Let $X_1, \ldots, X_n$ be iid Uniform$(0,\theta)$. It holds that

(f) $Q_n(\theta) \neq \bar{Q}_n(\theta)$, $\bar{Q}_n(\theta) < 0$, $\frac{d}{d\theta}E_\theta\{1\}$ cannot be differentiated under the integral.

Proof of CRBound. Well, the score function $\dot{l}_n(\theta; \underline{X})$ has mean 0, since

(a) $0 = \frac{d}{d\theta} E_\theta[1] = \frac{d}{d\theta} \int\cdots\int f_\theta(\underline{x})\, d\underline{x} \underset{\text{by (5)}}{=\!=} \int\cdots\int \frac{\partial}{\partial\theta} f_\theta(\underline{x})\, d\underline{x}$

(b) $= \int\cdots\int \frac{\frac{\partial}{\partial\theta} f_\theta(\underline{x})}{f_\theta(\underline{x})} f_\theta(\underline{x})\, d\underline{x} = \int\cdots\int \left[\frac{\partial}{\partial\theta} \log f_\theta(\underline{x})\right] f_\theta(\underline{x})\, d\underline{x}$

$= E_\theta\left\{\frac{\partial}{\partial\theta} \log f_\theta(\underline{X})\right\} = E_\theta\{\dot{l}_n(\theta; \underline{X})\}$.

Fact: If $EU = 0$, then $\mathrm{Cov}[U, V] = E\{(U-0)(V-\mu_V)\} = E[UV]$.

Thus

$$\mathrm{Cov}_\theta[T, \dot{l}_n(\theta; \underline{X})] = E_\theta\{T \cdot \dot{l}_n(\theta; \underline{X})\} = E_\theta\left\{T \cdot \frac{\partial}{\partial\theta} \log f_\theta(\underline{X})\right\}$$

$$= \int\cdots\int T(\underline{x}) \cdot \frac{\frac{\partial}{\partial\theta} f_\theta(\underline{x})}{f_\theta(\underline{x})} f_\theta(\underline{x})\, d\underline{x} = \int\cdots\int T(\underline{x}) \cdot \frac{\partial}{\partial\theta} f_\theta(\underline{x})\, d\underline{x}$$

$$\underset{\text{by (5)}}{=\!=} \frac{\partial}{\partial\theta} \int\cdots\int T(\underline{x}) f_\theta(\underline{x})\, d\underline{x} = \frac{d}{d\theta} E_\theta[T] = \frac{d}{d\theta} \tau(\theta)$$

(c) $= \tau'(\theta)$.

Fact: (Cauchy-Schwarz Inequality) $\left[\int\cdots\int g(\underline{x})\, h(\underline{x})\, d\underline{x}\right]^2 \le \left[\int\cdots\int g^2(\underline{x}) d\underline{x}\right] \cdot \left[\int\cdots\int h^2(\underline{x}) d\underline{x}\right]^2$ with equality if and only if $f(\underline{x}) = d\, g(\underline{x})$ for some constant with $d \neq 0$.

Applying Cauchy-Schwarz to (c) gives

(d) $[\tau'(\theta)]^2 = \{\mathrm{Cov}_\theta[T, \dot{l}_n(\theta; \underline{X})]\}^2 \le \mathrm{Var}_\theta[T]\, E_\theta\{[\dot{l}_n(\theta; \underline{X})]^2\} = \mathrm{Var}_\theta[T]\, \mathcal{I}_n(\theta)$,

with equality (for a fixed θ) if and only if $\dot{l}_n(\theta; \underline{X}) = A_n(\theta)[T(\underline{X}) - \tau(\theta)]$.

When this equality holds, we have

(e) $\mathcal{I}_n(\theta) = E_\theta\{[\dot{l}_n(\theta; \underline{X})]^2\} = A_n^2(\theta)\, E_\theta\{[T(\underline{X}) - \tau(\theta)]^2\} = A_n^2(\theta)\, \mathrm{Var}_\theta[T]$;

and (8) follows. HW 7.39 is to show that $\mathcal{I}_n(\theta) = \bar{\mathcal{I}}_n(\theta)$. □

Comment. If $T(\underline{X})$ is a biased estimator of $\tau(\theta)$, with bias $B(\theta) \equiv E_\theta T - \tau(\theta)$, then

$$\mathrm{Var}_\theta[T(\underline{X})] \ge \frac{[\dot{\tau}(\theta) + \dot{B}ias]^2}{\mathcal{I}_n(\theta)},$$

and this has been used to some effect (in a few cases) to minimize the MSE.

"PLANNED" Syllabus for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29	L1...Ch1: Probability, conditional probability, independence, and rvs.
Mar 31	L2...Sec2.1–2.3$^-$: E(·), mean, variance, and (especially) transformations (no mgf)
Apr 01	Q....HW and some catchup lecturing
	L3...Sampling with and without replacement
.....	
Apr 05	L4...Sec3.1–3.2(Handout): Bernoulli trials; Geometric, NegBinomialT (and Hypergeometric and Discrete Uniform)
	L5...Sec3.3(Handout): Poisson process, Poisson, Exponential, Gamma (and Beta and Uniform)
	Q ....HW, and finish Poisson process
	L6...Sec2.3: Moment generating function (mgf)
.....	
Apr 12	L7...More mgf, and CLT pics
	L8...Ch4: Joint, marginal, and conditional distributions
	Q....Sec4.1: Conditionals and convolution
	L9...Sec4.2: Conditional mean and variance
.....	
Apr 19	L10..Sec4.2 Independence and Expectation; Beta densities
	L11..Sec3.4 Exponential families; list of variance and covariance formulas
	Q....MIDTERM I
	L12..Discuss Midtem; the intuitive meaning of covariance; more on the var/cov list
.....	
Apr 26	L13..Sec4.5: Correlation inequaliaty; the empirical df; finish the var/cov list
	L14..Sec5.1–5.2: Properties of $\bar{X}$ and S^2
	Q....
Apr 30	L15..Sec4.5: Bivariate Normal Distribution
.....	
May 03	L16..Sec5.3: Derivation of Student-T and Fisher/Snedecor/varianceRatio-F
	L17..Handout: Moments via matrices
	Q....
	L18..Sec7.1–7.2:Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)
.....	
May 10	L19..Estimation: Method of Moments (MME) and Maximum Likelihood (MLE)MLE: Three illustrative MLE examples (Did Bernoulli(θ) and Uniform$(0,\theta)$).
	L20..More MLE.
	Q....Student-T technology.
	L21..MLEs for Gamma(α,β) and various Normal$(\mu(\theta),\sigma^2)$ examples
.....	
May 17	L22..MIDTERM II
	L23..Discuss Midterm
	Q....Bayes estimators
	L24..Bayes summary; begin Cramér–Rao inequality
.....	
May 24	L25..Bayes homework hints; Crameŕ–Rao inequality
	L26..UMVUEs, with examples
	Q....Asymptotic normalty of MLEs
	L27..Elementary ideas re hypothesis tests
.....	
May 31	MEMORIAL DAY HOLIDAY
June02	L28..Likelihood ratio tests
	Q....Work examples of LR-tests
June04	L29..Location and scale models (likely, the ten minute version)
.....	
June09	FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....	

Planned Homework for Stat 506 Spring 2004

by Galen R. Shorack

Mar 29 L1...Chapter1: #[2,5,11,13,14,25,33,38; do not be formal; should be easy];#17,18,21,23,24
Mar 31 L2...Chapter1: #44,53; Chapter2: #4,8c,17,19,24
Apr 01 Q....
L3...Chapter2: #20; Chapter3: #4,6,12,23,24b; Handout: #1
.....
Apr 05 L4...Handout: #2-5
L5...Handout: #6-7
Q ...
L6...
.....
Apr 12 L7...Chapter2: #9,16, (read 11); Chapter3: #22cd
L8...Chapter2: #33abc; Chapter3: #19,22ab
Q....
L9...Chapter2: # 26ab, 28ab, 30ad; Chapter3: #24c
.....
Apr 19 L10..Chapter4 #13,15; HW #1abc in Conditional Epectation handout about $v^2(y)$
L11..Unoficial: Work some off the HW sheet containing nine
Q....MIDTERM
L12..HW on the back of the midterm
.....
Apr 26 L13..Chapter4 #39, 43, 58
L14..Chapter4 #23(long, re transformations-start early), 44
Q....
Apr 30 L15..Chapter4 # 27, 30, 45
.....
May 03 L16..Chapter4 #19, 20(discuss #21 as well as a certain in class example in this context), 54
L17..Chapter4 #42, 52
Q....Covariance handout #1,2,4,6
L18..MLE handout #B2,B3
.....
May 10 L19..MLE handout #B1,B5
L20..MLE handout #B6,B7,B8
Q....
L21..
.....
May 17 L22..MIDTERM II
L23..Exercises on the back of Midterm II
Q....Chapter7 #24,23;
Find the Bayes estimators of $\theta(1-\theta)$ and $\theta/(1-\theta)$ in the Bernoulli(θ) Example 1.
L24..Chapter7 #22
.....
May 24 L25..None
L26..Chapter7 #47, 55ab, 19
Q....Evaluate the variance in $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, I^{-1}(\theta))$
in the context of #B3, #B5, #B8 on the earlier MLE homework handout
L27..#1 and #2 on the new homework handout
.....
May 31 MEMORIAL DAY HOLIDAY
June02 L28 #3 and #4 on the new homework handout.
Q
June04 L29..
.....
June09 FINAL EXAM(in our regular classroom). Wednesday June 9 at 8:30 - 10:20
.....

Stat 506 Homework

Let $X_1, \ldots, X_n$ be iid Poisson(θ)

#1 (a) Determine a reasonable test of $H_0: \theta \in \Theta_0$ vs $H_a: \theta > \theta_a$
Evaluate the rejection constant.
(b) Determine the power function.

#3. Determine the LR-test of $H_0: \theta = \theta_0$ vs $H_a: \theta \neq \theta_0$.

Let $X_1, \ldots, X_m$ be iid $N(\mu, \sigma^2)$.
Let $Y_1, \ldots, Y_n$ be iid $N(\nu, \sigma^2)$.

#2. (a) Determine a reasonable test of $H_0: \mu \leq \nu$ vs $H_a: \mu > \nu$.
Evaluate the rejection constant.

(b) Alter the test slightly, now supposing σ^2 is actually a known value σ_0^2.
(c) Determine the power function of the case (b) test.
(It will only depend on μ, ν and σ_0 thru $(\mu - \nu)/\sigma_0$.)

#4. Determine the LR-test of $H_0: \mu = \nu$ vs $H_a: \mu \neq \nu$.

SUFFICIENCY

DEF. 1. Let $f_\theta(\cdot)$ (with $\theta \in \Theta$) be a family of densities of $\underset{\sim}{X}$. Let $T \equiv T(\underset{\sim}{X})$ denote some function of $\underset{\sim}{X}$. We say that T is *sufficient* for θ if the conditional distribution

(1) $f_{\underset{\sim}{X}|T=t}(\cdot)$ does not depend on θ.

Note the implications that:

(2) $f_\theta(\underset{\sim}{x}) = g_\theta(t)\, f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x})$ — We can reconstruct natures $\underset{\sim}{X}$ experiment from natures T-experiment and "random number table". We have a factorization.

(3) $f_\theta(\underset{\sim}{x}) = g_\theta(t)\, h_t(\underset{\sim}{x})$

THEOREM 1 (Factorization theorem) T is sufficient for θ iff we can write

(4) $f_\theta(\underset{\sim}{x}) = g_\theta(t)\, h_t(\underset{\sim}{x})$.

Proof. From above, sufficiency gives a factorization.

So, suppose we have a factorization. Then the marginal distribution of T:

(a) $$f_{\theta,T}(t) = \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} f_\theta(\underset{\sim}{x}) = \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} g_\theta(t)\, h(\underset{\sim}{x}) = g_\theta(t) \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x}).$$

And so

(b) $$f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } T(\underset{\sim}{x}) \neq t \\ \dfrac{f_\theta(\underset{\sim}{x})}{f_{\theta,T}(t)} = \dfrac{g_\theta(t)\, h(\underset{\sim}{x})}{g_\theta(t) \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x})} = \dfrac{h(\underset{\sim}{x})}{\sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x})} & \end{cases}$$

(c) = (an expression that does not depend on θ).

Thus T is sufficient for θ. □

EXAMPLE 1 (Bernoulli(θ) trials) Let $X_1, \ldots, X_n$ be iid Bernoulli(θ). Let $T \equiv \sum_{i=1}^n X_i$. Then

$$f_\theta(\underset{\sim}{x}) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$= \theta^t (1-\theta)^{n-t} \cdot 1 \qquad (= \underline{\text{a}} \text{ factorization}, \therefore \text{ sufficiency})$$

(a) $$= \binom{n}{t}\theta^t(1-\theta)^{n-t} \cdot \frac{1}{\binom{n}{t}} \qquad (= \text{marginal} \times \text{conditional})$$

Let's reconstruct the $\underset{\sim}{X}$ experiment. The fact that such a reconstruction is possible, motivates the definition. □

DEF 2. A statistic $S \equiv S(\underset{\sim}{X})$, in the above model, is called *ancillary* if the distribution of S does not depend on θ.

COMPLETENESS

T function of data

DEF 1 A family $\mathcal{P}$ of distributions of T is called *complete* if

(1) $0 = E_{F_T}\, g(T) = \int g(t) f_T(t)\,dt$ for all distributions $f_T(t)$ in $\mathcal{P}$ implies that $g(t) = 0$ for "all" t.

(any function of T)

This a mathematical condition, not a statistical condition.

Math Fact 1. If $h(x) \equiv \sum_{k=0}^{\infty} a_k x^k = 0$ for all x in an interval, then all $a_k = 0$.

Math Fact 2. If $\int_{(-\infty, x]} h(t)\,dt = 0$ for all x, then $h(x) = 0$ for "all" x.

"Ff": First, $\int_{(x,y]} h(t)\,dt = 0$. Then $\int_A h(t)\,dt = 0$ for "all" A. Thus $h(x) = 0$ for "all" x. □

Example 1. (Poisson(θ)) Let $X_1, \ldots, X_n$ be iid Poisson(θ). Then

$$f_\theta(\underline{x}) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-\theta n}\theta^{\sum x_i}}{\prod_{i=1}^n x_i!} = \left\{e^{-\theta n}\theta^{t}\right\}\left\{\frac{1}{\prod_{i=1}^n x_i!}\right\}$$

$$= \{g_\theta(t)\}\ \{h(\underline{x})\}$$

for $T \equiv T(\underline{X}) \equiv \sum_{i=1}^n X_i$. Thus T is sufficient for θ. And:

(2) The family of distributions $f_\theta(t)$ of T is complete.
The family of all Poisson($n\theta$) distributions, with $\theta > 0$, is complete.
The family of all Poisson(θ) distributions, with $\theta > 0$, is complete.

Why? Suppose $0 = E_\theta\{g(T)\}$ for all $\theta > 0$. Then

(a) $$0 = E_\theta\{g(T)\} = \sum_{k=0}^{\infty} g(k) e^{-\theta}\theta^k/k! = e^{-\theta}\sum_{k=0}^{\infty}\left(\frac{g(k)}{k!}\right)\theta^k \quad \text{for all } \theta;$$

Since $e^{-\theta}$ is never 0, this means that

(b) $$0 = \sum_{k=0}^{\infty} \frac{g(k)}{k!}\theta^k = \sum_{k=0}^{\infty} a_k \theta^k \quad (\text{with } a_k \equiv \frac{g(k)}{k!}) \quad \text{for all } \theta;$$

and thus $a_k = 0$ for all k by Math Fact 1. Thus $g(k) = 0$ for all k.
Thus (2) is true; i.e. the Poisson(θ), $\theta > 0$, family is complete. □

Example 2. (Uniform$(0,\theta)$). Suppose $X \equiv$ Uniform$(0,\theta)$. (i) Suppose

$$0 = E_\theta\{g(X)\} = \int_0^\theta g(x)\frac{1}{\theta}\,dx \quad \text{for all } \theta.$$

Then $0 = \int_0^\theta g(x)\,dx$ for all θ. Thus $g(x) = 0$ for "all" x. Thus:

(3) The family of all Uniform$(0,\theta)$, $\theta > 0$, distributions is complete.

(ii) Now $T \equiv \max_{1 \le i \le n} X_i$ is sufficient for a sample $X_1, \ldots, X_n$ from Uniform$(0,\theta)$.
And

$$0 = E_\theta\{g(T)\} = \int_0^\theta g(t)\frac{n t^{n-1}}{\theta^n}\,dt = \left\{\frac{n}{\theta^n}\right\}\int_0^\theta g(t) t^{n-1}\,dt \quad \text{for all } \theta.$$

Implies $g(t)t^{n-1} = 0$ for all t; which implies $g(t) = 0$ for "all" t; so T is complete. □

EXPONENTIAL FAMILY

GRS

Definition 1. (Exponential family) Suppose we have densities $f_\theta(\cdot)$, $\theta \in \Theta$, of the form

$$(1) \quad f_\theta(x) = f(x|\theta) = C(\theta)\, e^{\sum_{i=1}^{k} w_i(\theta)\, T_i(x)}\, h(x)$$

for function $T_1, \ldots, T_k$ of the observations, and functions $C(\theta)$, $w_1(\cdot), \ldots, w_k(\cdot)$ of the parameters. Then $\{f_\theta : \theta \in \Theta\}$ is a <u>exponential family</u>.

Definition 2 (Natural parameter) For $\eta_1(\theta), \ldots, \eta_k(\theta)$, we may reparameterize:

$$(2) \quad = C^*(\eta)\, e^{\sum_{i=1}^{k} \eta_i\, T_i(x)}\, h(x).$$

Note: The <u>natural parameter space</u> (values of η for which this is indeed a density) is <u>convex</u>, and $E_\eta\{g(X)\}$ is an <u>analytic function</u> on the interior, provided only it exists.

Example 1 [Bernoulli trials, $0 < p < 1$]

$$f_p(x) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

$$= (1-p)^n\, e^{(\sum x_i)\log\frac{p}{1-p}} \cdot 1$$

$$(a) \quad = C^*(\eta)\, e^{T(x)\,\eta} \cdot 1$$

where $\eta \equiv \eta(p) \equiv \log\frac{p}{1-p} \equiv$ (log odds). (Note: $p = \frac{1}{2}$ iff $\eta = 0$.)

[Comment: Binomial (n,p) has $h(x) = \binom{n}{x}$.]

Comment. $X \equiv (X_1, \ldots, X_n)'$ are iid Bernoulli (p); so $T \equiv T(X) \cong$ Binomial(n,p).

Then $f_{X|T=t}(x) = \begin{cases} \frac{p^t(1-p)^t}{\binom{n}{t}p^t(1-p)^t} = \frac{1}{\binom{n}{t}} & \text{for all } x \text{ such that } \sum x_i = t \\ 0 & \text{else.} \end{cases}$

Thus the conditional distribution of X given that $T \equiv T(X) = t$ does not depend on p. That is, T is sufficient for p in this model. □

Example 2. [Poisson (θ)] $f_\theta(x) = e^{-\theta}\frac{\theta^x}{x!} = \{e^{-\theta}\}\, e^{x(\log\theta)}\left\{\frac{1}{x!}\right\}$, for $x = 0, 1, 2, \ldots$.
Then $T \equiv \sum X_i$ is sufficient (do it 2 ways); re an iid sample. □

Example 3. [Normal (μ, σ^2), with $\theta \equiv (\mu, \sigma^2)'$].

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \left\{\frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{\mu^2}{2\sigma^2}}\right\} e^{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2}$$

If $X_1, \ldots, X_n$ is an iid sample, then $T = (T_1, T_2)' = (\sum X_i, \sum X_i^2)'$ on $(-\infty, \infty)$.
is sufficient for θ. □

Sufficiency and Completeness for Exponential Families

Suppose now that $X_1, \ldots, X_n$ are a sample from the exponential family

(3) $f_\theta(x) = C(\theta)\, e^{\sum_{i=1}^{k} w_i(\theta) T_i(x)}\, h(x)$.

For $1 \le i \le k$, let $T_i \equiv T_i(\underset{\sim}{X}) \equiv \sum_{j=1}^{n} T_i(X_j)$; and let $T \equiv T(\underset{\sim}{X}) \equiv (T_1, \ldots, T_k)$.

(4) T is sufficient for θ

Proof (Factorization) Well,

$$f_\theta(\underset{\sim}{x}) = \prod_{j=1}^{n} f_\theta(x_j) = \left\{ C^n(\theta)\, e^{w_1(\theta)\sum_{j=1}^{n} T_1(x_j) + \cdots + w_k(\theta)\sum_{j=1}^{n} T_k(x_j)} \right\} \left\{ \prod_{j=1}^{n} h(x_j) \right\}$$

(a) $= g_\theta(t_1, \ldots, t_k) \cdot h(\underset{\sim}{x})$.

(Conditional distribution) [Let $k=1$.] Let $T \equiv T(\underset{\sim}{X}) \equiv \sum_{j=1}^{n} T(X_j)$. Then

$g_\theta(t) \equiv P_\theta(T=t) = \sum_{\{\underset{\sim}{x}:\, T(\underset{\sim}{x})=t\}} P_\theta(\underset{\sim}{X}=\underset{\sim}{x})$ (in the discrete case)

$$= \sum_{\{\underset{\sim}{x}:\, T(\underset{\sim}{x})=t\}} C^n(\theta)\, e^{w(\theta)\, T(\underset{\sim}{x})}\, h(\underset{\sim}{x}) = C^n(\theta) e^{w(\theta)t} \underbrace{\sum_{\{\underset{\sim}{x}:\, T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x})}_{\equiv h^*(t)}$$

(b) $= C^n(\theta)\, e^{w(\theta)t}\, h^*(t) =$ (the Exponential family of the dist of T)

and thus

(c) $$P_{X|T=t}(\underset{\sim}{x}) = \frac{C^n(\theta)\, e^{t\, w(\theta)} \prod_{j=1}^{n} h(x_j)}{C^n(\theta)\, e^{t\, w(\theta)}\, h^*(t)} = \frac{\prod_{j=1}^{n} h(x_j)}{h^*(t)} = \left(\text{does not depend on } \theta\right)$$

Thus T is sufficient for θ. □

Note from (b) that the sufficient statistic $T \equiv (T_1, \ldots, T_k)'$ is such that

(5) T is distributed as an exponential family.

[The conditional distribution of $T_1 | T_2 = t_2$ is also a particular kind of exponential family.]

We now turn to completeness: Let θ range over the parameter space Θ. Then:

(6) The family of distributions of the sufficient statistic is complete if $(w_1(\theta), \ldots, w_k(\theta))$ traces out a region containing a k-dimensional rectangle, for $\theta \in \Theta$.

Example 4 Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$. (a) Then

$$f_{\mu,\sigma^2}(\underset{\sim}{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}}\, e^{\frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2}.$$

Since $\left\{\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right) : -\infty < \mu < \infty, \sigma^2 > 0\right\}$ contains a 2-dimensional rectangle, the family of distributions of $\left(\sum_{1}^{n} X_i, \sum_{1}^{n} X_i^2\right)$ is complete.

(b) Let $\mu = \sigma^2 = \theta$. We only get a 1-dimensional curve. Also, $E_\theta(\bar{X} - S^2) = 0$ for all $\theta > 0$. So $N(\theta, \theta)$, with $\theta > 0$, is not complete. □

uniform unbiased 26-5-2004

UMVU ESTIMATORS

NOTATION 1. Consider a family of models $\underset{\sim}{X} \cong f_\theta(\cdot)$ for some θ in the parameter set. Suppose $T = T(\underset{\sim}{X})$ is sufficient for θ. Let $\tau(\theta)$ be of interest, and suppose $W \equiv W(\underset{\sim}{X})$ is such that trying to estimate $\tau(\theta)$

(1) $E_\theta[W] = \tau(\theta)$, so that W is unbiased for $\tau(\theta)$.

Math Fact. $E[Y] = E[m(X)]$ and $\mathrm{Var}[Y] = \mathrm{Var}[m(X)] + E[v(X)]$

THEOREM 1. (Rao-Blackwell) Let $V \equiv V(T) \equiv E(W|T)$; note that V does not depend on θ since T is sufficient for θ. Moreover:

(2) $E_\theta[V] = E_\theta\{E[W|T]\} = E_\theta[W] = \tau(\theta)$; so V is unbiassed for $\tau(\theta)$.

(3) $\mathrm{Var}_\theta[V] = \mathrm{Var}_\theta[E[W|T]] \le \mathrm{Var}_\theta[E[W|T]] + E_\theta[\mathrm{Var}[W|T]] = \mathrm{Var}_\theta[W]$,

so that V is at least as good as W among such unbiased estimators.

THEOREM 2. (Lehmann-Scheffe') If T is both sufficient and complete, then

(4) V is the unique UMVUE for $\tau(\theta)$.

[UMVUE for $\tau(\theta)$ means it is the uniform minimum variance unbiased estimator.]

Proof. Let U denote any other unbiased estimator of $\tau(\theta)$. Let $\bar{U} \equiv E(U|T)$. Then $h(T) \equiv \bar{U} - V$ is a function of T having $E_\theta[h(T)] = 0$ for all θ. But T is complete; so $\bar{U} - V = h(T) = 0$, or $\bar{U} = V$. Aha! There is only one unbiased estimator of $\tau(\theta)$ that is a function of T.

That is $E(U|T) = \bar{U} = V = E(W|T)$ as for any unbiased estimators U and W; and this single $\bar{U} = V$ is at least as good as W (and U)

smart or stupid guess (unbiased) give same result because of completeness

EXAMPLE 1. (Bernoulli(θ)) Let $X_1, \ldots, X_n$ be iid Bernoulli(θ). Then $T = \sum_{i=1}^n X_i$ is sufficient for θ, and it is comptete. Since $W = X_1$ is an unbiased estimator of θ, $V \equiv E[W|T] = E[X_1 | \sum_{i=1}^n X_i]$ is the UMVUE of θ. Here,

(4) $V(t) \equiv E[X_1 | T = t] = \frac{t}{n}$, (i.e. $V = \frac{T}{n} = \bar{X}_n$)

since knowing t of the n values are 1 and the rest are 0 means that X_1 has a $\frac{t}{n}$ chance of being a 1. □

THEOREM 3. If V is an UMVUE of $\tau(\theta)$, then V is unique.

Proof. Let V' denote a second UMVUE estimator. Set $V^* \equiv \frac{1}{2}(V+V')$. Then

$$\mathrm{Var}_\theta[V^*] = \tfrac{1}{4}\mathrm{Var}_\theta[V] + \tfrac{1}{4}\mathrm{Var}_\theta[V'] + \tfrac{1}{2}\mathrm{Cov}_\theta[V,V'] \quad \left\{\begin{array}{l}\text{with } \mathrm{Var}_\theta[V'] = \mathrm{Var}_\theta[V] \\ \text{as } V' \text{ is also UMVUE}\end{array}\right\}$$

(a) $$\leq \mathrm{Var}_\theta[V] \qquad \begin{array}{l}\text{by the Correlation Inequality} \\ \text{as } |\mathrm{Cov}_\theta[V,V']|^2 \leq \mathrm{Var}_\theta[V]\,\mathrm{Var}_\theta[V'] \leq \mathrm{Var}_\theta^2[V]\end{array}$$

with $<$ unless $V' = aV$ (with $a=1$, by unbiased) a.s. □

REMARK 1. Similarly, if V is an UMVUE of $\tau(\theta)$ and $E_\theta U = 0$ for all θ, $V+\lambda U$ is also unbiased for $\tau(\theta)$ (for every real λ). Moreover,

$$\mathrm{Var}_\theta[V+\lambda U] = \mathrm{Var}_\theta[V] + 2\lambda\,\mathrm{Cov}_\theta[U,V] + \lambda^2\,\mathrm{Var}_\theta[U]$$

$$< \mathrm{Var}_\theta[V] \quad \text{for } \lambda \text{ between } 0 \text{ and } \frac{2\,\mathrm{Cov}_\theta[U,V]}{\mathrm{Var}_\theta[U]}$$

(just think of the parabola in λ, and when $\mathrm{Var}_\theta[V]$ is the bottom point).

Thus $U+$(someλ)V is better than V for some θ, unless $\mathrm{Cov}_\theta[U,V] = 0$ for all θ. Thus

(5) $\mathrm{Cov}_\theta[U,V] = 0$ whenever V is UMVUE for $\tau(\theta)$ and U is unbiased for 0. □

THEOREM 4. Let V be unbiased for $\tau(\theta)$. Then

(6) V is UMVUE for θ iff $\mathrm{Cov}_\theta[U,V] = 0$ for all U unbiased for 0

Proof. This is an immediate corollary to Remark 1 and the proof method of Theorem 3. □

Asymptotic Properties of MLEs

Notation. Let $X_1, \dots, X_n$ be iid $f_\theta(\cdot)$ for some unknown value of the parameter. The observations were taken for the purpose of estimating the true parameter value. The family of densities $f_\theta(\cdot)$ is "sufficiently nice".

Preliminary Facts. Then the score function $\dot{l}_n(\theta; \underset{\sim}{X})$ satisfies

$$\frac{1}{n}\dot{l}(\theta; \underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^n \dot{l}(\theta; X_i) = \frac{1}{n}\sum_{i=1}^n \underbrace{\frac{\partial}{\partial\theta}\log f_\theta(X_i)}_{\equiv Y_i \overset{\cdot}{\sim}(0, \mathcal{I}(\theta))} = \sqrt{n}(\bar{Y}_n - \mu_Y)$$

(1) converges to $\to_d N(0, \sigma_Y^2) = N(0, \mathcal{I}(\theta))$, by the CLT.

Also,

$$I_n \equiv \frac{1}{n}\ddot{l}(\theta; \underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^n \ddot{l}(\theta; X_i) = \frac{1}{n}\sum_{i=1}^n \underbrace{-\frac{\partial^2}{\partial\theta^2}\log f_\theta(X_i)}_{\equiv V_i} = \bar{V}_n$$

(2) $\to_p \mu_V = E_\theta\left\{-\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right\} = \bar{\mathcal{I}}(\theta) = \mathcal{I}(\theta)$ by the WLLN.

Theorem 1. The MLE $\hat{\theta}_n$ and the observed information $\hat{I}_n \equiv \frac{1}{n}\ddot{l}(\hat{\theta}_n; \underset{\sim}{X})$ satisfy:

(3) $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \mathcal{I}^{-1}(\theta))$.

(4) $\hat{I}_n \to_p \mathcal{I}(\theta)$.

(5) $\sqrt{n}(\hat{\theta}_n - \theta)\sqrt{\hat{I}_n} \to_d N(0, 1)$.

Thus, an asymptotically correct confidence interval for θ is given by

(6) $\hat{\theta}_n \pm z_{\frac{\alpha}{2}} / \sqrt{n\hat{I}_n}$.

[Use of $\hat{I}_n$ is typically more accurate than use of $\mathcal{I}(\hat{\theta}_n)$.]

Law of large numbers

$$P(|\bar{X} - \mu| > \epsilon) = P(|\bar{X} - \mu|^2 > \epsilon^2) \le \frac{E|\bar{X}_n - \mu|^2}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2}$$

If n gets big $|\bar{X} - \mu|$ becomes small, even if σ^2 big. Doesn't tell how big n needs to be though.

Proof. The MLE satisfies $0 = \dot{l}(\hat{\theta}_n; \underset{\sim}{X}) = \dot{l}(\theta; \underset{\sim}{X}) + (\hat{\theta}_n - \theta)\ddot{l}(\theta_n^*; \underset{\sim}{X})$ by the mean value thm. ($x - x_0$; one term Taylor series; mean value theorem)

Thus

(a) $$0 = \frac{1}{\sqrt{n}}\dot{l}(\hat{\theta}_n; \underset{\sim}{X}) = \underbrace{\frac{1}{\sqrt{n}}\dot{l}(\theta; \underset{\sim}{X})}_{\to_d N(0, \mathcal{I}(\theta)) \text{ by (1)}} - \sqrt{n}(\hat{\theta}_n - \theta) \cdot \underbrace{\frac{1}{n}\ddot{l}(\theta_n^*; \underset{\sim}{X})}_{\to_p \mathcal{I}(\theta) \text{ "as in" (2)}}$$

(divide by $\sqrt{n}$)

Thus $$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\dot{l}(\theta; \underset{\sim}{X}) / I_n^* \to_d N(0, \mathcal{I}(\theta))/\mathcal{I}(\theta) = N(0, \mathcal{I}^{-1}(\theta))$$

by Slutsky's theorem. □

(Rewrite this for vectors. How? matrix)

$$Y_i = \dot{\ell}(\theta; X_i) = \frac{\partial}{\partial\theta} \log f_\theta(X_i)$$

μ Information

$$\cong (0, \sigma_Y^2 = E_\theta[\quad''\quad]^2 = \mathcal{I}(\theta))$$

$$\frac{1}{\sqrt{n}} \sum_1^{\hat{}} \overline{\dot{\ell}(\theta, X_i)} = \boxed{\sqrt{n}(\bar{Y} - 0)} \to_{\text{dist}} N(0, \sigma_Y^2) = N(0, \mathcal{I}(\theta)) \quad \text{CLT}$$

converges to N

$$V_i = -\ddot{\ell}(\theta; X_i) = -\frac{\partial^2}{\partial\theta^2} \log f_\theta(X_i) \cong (\bar{\mathcal{I}}(\theta), \boxtimes)$$

$$\text{So } \bar{V} \xrightarrow[P]{} m_V = \mathcal{I}(\theta) \underset{\text{rng}}{=} \mathcal{I}(\theta) \quad \text{WLL}$$

stacks at point probability P

$$\frac{\sqrt{n}(\bar{Y} - \theta)}{\bar{V}} = \frac{1}{\mathcal{I}(\theta)} N(0, \mathcal{I}(\theta)) = N(0, \overline{\mathcal{I}(\theta)})$$

$$\sqrt{n}(\hat{\theta} - \theta) \to N(0, \mathcal{I}(\theta)^{-1})$$

$\left\{\left(\frac{\mu+\nu}{\sigma^2}, \frac{2\mu-\nu}{\sigma^2}, -\frac{1}{2\sigma^2}\right): \sigma^2>0, -\infty<\mu,\nu<\infty\right\}$ 3D rect.

Thus (T_1, T_2, T_3) is suff for $\theta \equiv (\mu, \nu, \sigma^2)$ 3 coord vector is suff for 3 coord vector.

Ex Let $X_1, .., X_m$ be iid $N(\mu+\nu, \sigma^2)$ and $Y_1, .., Y_n$ be iid $N(2\mu-\nu, k\sigma^2)$; these independent.

$$f = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} e^{-\frac{1}{2\sigma^2}\sum_1^m (x_i-(\mu+\nu))^2} \left(\frac{1}{2\pi k\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2k\sigma^2}\sum_1^n (y_j-(2\mu-\nu))^2}$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{m}{2}}\left(\frac{1}{2\pi k}\right)^{\frac{n}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{m+n}{2}} e^{-\frac{1}{2\sigma^2}\left[\sum_1^m x_i^2 + \frac{1}{k}\sum_1^n y_j^2\right] + \frac{\mu+\nu}{\sigma^2}\sum_1^m x_i + \frac{2\mu-\nu}{2\sigma^2}\sum_1^n y_j} e^{-\frac{m(\mu+\nu)^2}{2\sigma^2} - \frac{n(2\mu-\nu)^2}{2k\sigma^2}}$$

Thus

$T \equiv \left(\sum_1^m x_i, \sum_1^n y_j, \left(\sum_1^m x_i^2 + \frac{1}{k}\sum_1^n y_j^2\right)\right)$ is sufficient and complete for $\Theta \equiv (\mu, \nu, \sigma^2)$.

So is $\left(\bar{X}_m, \bar{Y}_n, \left(\sum_1^m x_i^2 + \frac{1}{k}\sum_1^n y_j^2\right) - m\bar{X}_m^2 - \frac{n}{k}\bar{Y}_n^2\right)$

$$= \left(\sum_1^m x_i^2 - m\bar{X}_n^2\right) + \frac{1}{k}\left(\sum_1^n y_i^2 - n\bar{Y}_n^2\right)$$

$$= SS_{xx} + \frac{1}{k} SS_{yy}$$

So is $T \equiv \left(\bar{X}_m, \bar{Y}_n, S_p^2 \equiv \frac{SS_{xx} + \frac{1}{k} SS_{yy}}{m+n-2}\right)$. $\left\{\text{The MLE is } \hat{\sigma}^2 \equiv \frac{SS_{xx} + \frac{1}{k} SS_{yy}}{m+n}\right\}$

Since $\left\{\left(\frac{\mu+\nu}{\sigma^2}, \frac{2\mu-\nu}{\sigma^2}, -\frac{1}{2\sigma^2}\right): \sigma>0, -\infty<\mu,\nu<\infty\right\}$ contains a 3-dimensional rectangle,

the family of distributions of T' is complete.

Fact. Since the $T' \equiv T'(\underline{x}, \underline{y}) \longrightarrow T(\underline{x}, \underline{y}) \equiv T$ mapping is 1-1,

(a) the family of distributions of T is sufficient and complete.

Consider any function $\tau(\theta) \equiv \tau(\mu, \nu, \sigma^2)$.

Let $V \equiv V(T)$ denote some function of T.

Fact If

(*) $E_\theta V = \tau(\theta)$ for all θ, then V is the ! UMVUE of $g(\theta)$.

Thus

$\bar{X}$ is the ! UMVUE of $\mu+\nu$

$\bar{Y}$ is the ! UMVUE of $2\mu-\nu$

$\hat{\mu} \equiv \frac{\bar{X}+\bar{Y}}{3}$ is the ! UMVUE of μ.

$\hat{\nu} \equiv \frac{2\bar{X}-\bar{Y}}{3}$ is the ! UMVUE of ν

S_p^2 is the ! UMVUE of σ^2

$c_{m,n} S_p$ is the ! UMVUE of σ (for some constant $c_{m,n} \to 1$ as $m \wedge n \to \infty$).

Case When σ^2 is Unknown

11. A bakery was taken to court for selling loaves of bread that were underweight. These loaves were advertised as weighing 24 ounces. In its defense, the bakery claimed that the advertised weight was meant to imply not that each loaf weighed exactly 24 ounces, but rather that the average value over all loaves was 24 ounces. The prosecution in a rebuttal produced evidence that a randomly chosen sample of 20 loaves had an average weight of 22.8 ounces with a sample standard deviation of 1.4 ounces. In her ruling, the judge stated that advertising a weight of 24 ounces would be acceptable if the mean weight were at least 23 ounces.
    (a) What hypothesis should be tested?
    (b) For the 5 percent level of significance, what should the judge rule?

13. A car is advertised as getting at least 31 miles per gallon in highway driving on trips of at least 100 miles. Suppose the miles per gallon obtained in 8 independent experiments (each consisting of a nonstop highway trip of 100 miles) are

    28, 29, 31, 27, 30, 35, 25, 29

    (a) If we want to check if these data disprove the advertising claim, what should we take as the null hypothesis?
    (b) What is the alternative hypothesis?
    (c) Is the claim disproved at the 5 percent level of significance?
    (d) What about at the 1 percent level?

14. A manufacturer claims that the mean lifetime of the batteries it produces is at least 250 hours of use. A sample of 20 batteries yielded the following data:

    237, 254, 255, 239, 244, 248, 252, 255, 233, 259, 236, 232, 243, 261, 255, 245, 248, 243, 238, 246

    (a) Are these data consistent, at the 5 percent level, with the claim of the manufacturer?
    (b) What about at the 1 percent level?

15. A water official insists that the average daily household water use in a certain county is at least 400 gallons. To check this claim, a random sample of 25 households was checked. The average of those sampled was 367 with a sample standard deviation of 62. Is this consistent with the official's claim?

17. A medical scientist believes that the average basal temperature of (outwardly) healthy individuals has increased over time and is now greater than 98.6°F (37°C). To prove this, she has randomly selected 100 healthy individuals. If their mean temperature is 98.74°F with a sample standard deviation of 1.1°F, does this prove her claim at the 5 percent level? What about at the 1 percent level?

18. In 1981, entering students at a certain university had an average score of 542 on the verbal part of the SAT. A random sample of the scores of 20 students in the 1993 class resulted in these scores:

    542, 490, 582, 511, 515, 564, 500, 602, 488, 512, 518, 522, 505, 569, 575, 515, 520, 528, 533, 515

    Do the above data prove that the average score has decreased to below 542? Use the 5 percent level of significance.

Case when σ_0^2 is Known

2. A previous sample of fish in Lake Michigan indicated that the mean polychlorinated biphenyl (PCB) concentration per fish was 11.2 parts per million with a standard deviation of 2 parts per million. Suppose a new random sample of 10 fish has the following concentrations:

   11.5, 12.0, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6

   Assume that the standard deviation has remained equal to 2 parts per million, and test the hypothesis that the mean PCB concentration has also remained unchanged at 11.2 parts per million. Use the 5 percent level of significance.

6. It is known that the value received at a local receiving station is equal to the value sent plus a random error that is normal with mean 0 and standard deviation 2. If the same value is sent 7 times, compute the p value for the test of the null hypothesis that the value sent is equal to 14, if the values received are

   14.6, 14.8, 15.1, 13.2, 12.4, 16.8, 16.3

7. Historical data indicate that household water use tends to be normally distributed with a mean of 360 gallons and a standard deviation of 40 gallons per day. To see if this is still the situation, a random sample of 200 households was chosen. The average daily water use in these households was then seen to equal 374 gallons per day.
   (a) Are these data consistent with the historical distribution? Use the 5 percent level of significance.
   (b) What is the p value?

8. When a certain production process is operating properly, it produces items that each have a measurable characteristic with mean 122 and standard deviation 9. However, occasionally the process goes out of control, and this results in a change in the mean of the items produced. Test the hypothesis that the process is presently in control if a random sample of 10 recently produced items had the following values:

   123, 120, 115, 125, 131, 127, 130, 118, 125, 128

   Specify the null and alternative hypotheses, and find the p value.

9. A leasing firm operates on the assumption that the annual number of miles driven in its leased cars is normally distributed with mean 13,500 and standard deviation 4000 miles. To see whether this assumption is valid, a random sample of 36 one-year-old cars has been checked. What conclusion can you draw if the average mileage on these 36 cars is 15,233?

11. Traffic authorities claim that traffic lights are red for a time that is normal with mean 30 seconds and standard deviation 1.4 seconds. To test this claim, a sample of 40 traffic lights was checked. If the average time of the 40 red lights observed was 32.2 seconds, can we conclude, at the 5 percent level of significance, that the authorities are incorrect? What about at the 1 percent level of significance?

12. The number of cases of childhood cancer occurring within a 3-year span in communities of a specified size approximately has a normal distribution with mean 16.2 and standard deviation 4.7. To see whether this distribution changes when the community is situated near a high-level electromagnetic field, a researcher chose such a community and subsequently discovered that there had been a total of 32 cases of childhood cancers within the last 3 years. Using these data, find the p value of the test of the hypothesis that the distribution of the number of childhood cancers in communities near high-level electromagnetic fields remains normal with mean 16.2 and standard deviation 4.7.

HYPOTHESIS TESTING

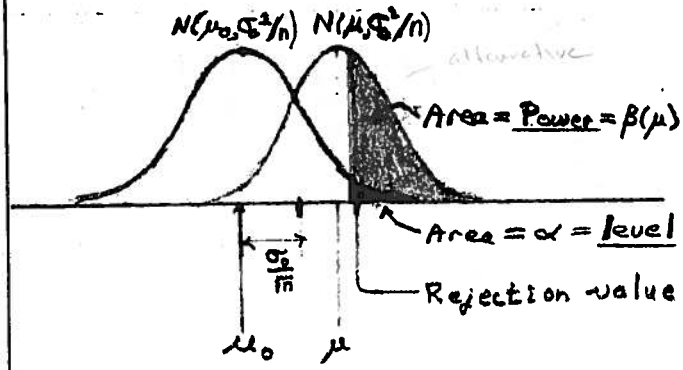
EXAMPLE 1. (Power, in the $N(\mu, \sigma_0^2)$ model)

Model our experiment as $X_1, \ldots, X_n$ iid $N(\mu, \sigma_0^2)$. We want:

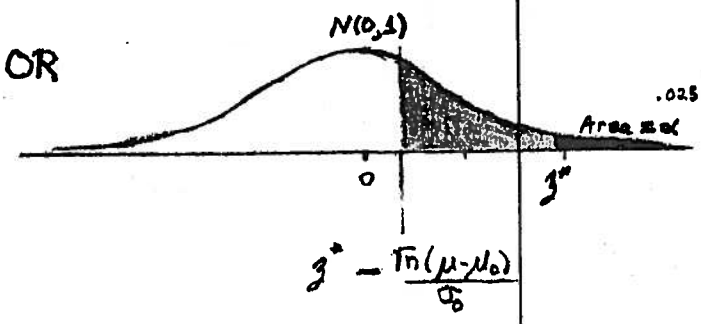
To test $H_0: \mu = \mu_0$ vs $H_1: \mu = \mu_1$ (where $\mu_1 > \mu_0$).

A reasonable <u>decision rule</u> is:

Reject H_0 in case $\bar{X}_n > \mu_0 + z^* \frac{\sigma_0}{\sqrt{n}}$ ≡ (the rejection value)



OR



The power function of this decision rule satisfies:

$$\text{Power}(\mu_0) = P(\text{Reject } H_0 \mid H_0 \text{ is true}) = P_{\mu_0}(\text{Reject } H_0)$$

$$= P_{\mu_0}\left(\bar{X}_n > \mu_0 + z^* \frac{\sigma_0}{\sqrt{n}}\right) = P_{\mu_0}\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_0} > z^*\right)$$

$$= P(N(0,1) > z^*) = \alpha \quad (\text{for } z^* \equiv z^{(\alpha)}),$$

and

$$\text{Power}(\mu) = P_\mu(\text{Reject } H_0) = P_\mu\left(\bar{X}_n > \mu_0 + z^* \frac{\sigma_0}{\sqrt{n}}\right)$$

$$= P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma_0} > z^* - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma_0}\right)$$

$$= P\left(N(0,1) > z^* - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\right) = P(N(0,1) > z^* - \delta)$$

where $\delta \equiv \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}$, and the power depends on n, μ, μ_0, σ_0 only through δ.

[This problem (with σ_0 known) has limited utility, but it is easy to do exact calculations.]

If the power is not good enough you must increase either α or n, or redesign a new experiment with smaller σ.

EXAMPLE 2. (Power, for Bernoulli(p) trials)

For $X_1, \ldots, X_n$ iid Bernoulli(p) we want:

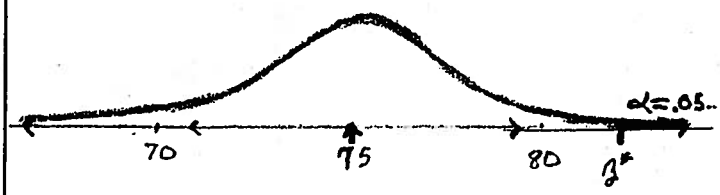
To test $H_0: p = p_0 (= .75, \text{say})$ vs $H_1: p > p_0$. (Perhaps, $n = 100$)

It seems reasonable to reject H_0 if $\hat{p}_n = \bar{X}_n$ (or $\sum_{i=1}^n X_i$) is "too big".

Now when H_0 is true, $\sum_1^n X_i \cong \text{Binomial}(n, p_0) = \text{Binomial}(100, .75) \doteq N(75, 4.33^2)$.

An reasonable decision rule is thus given by:

Reject H_0 in case $\sum_1^n X_i > np_0 + z^* \sqrt{np_0q_0} = 75 + 1.645 \times 4.33 = 82.12$



The power function of this decision rule satisfies:

$$\text{Power}(p_0) = P_{p_0}(\text{Reject } H_0) = P_{p_0}\left(\sum_1^n X_i > np_0 + z^*\sqrt{np_0q_0}\right)$$

$$\doteq P(N(0,1) > z^*) = \alpha \quad (\text{for } z^* = z^{(\alpha)}),$$

and

$$\text{Power}(p) = P_p(\text{Reject } H_0) = P_p\left(\sum_1^n X_i > np_0 + z^*\sqrt{np_0q_0}\right)$$

$$= P_p\left(\frac{\sum_1^n X_i - np}{\sqrt{npq}} > \frac{-np + np_0 + z^*\sqrt{np_0q_0}}{\sqrt{npq}}\right)$$

$$\doteq P\left(N(0,1) > z^*\sqrt{\frac{p_0q_0}{pq}} - \frac{\sqrt{n}(p-p_0)}{\sqrt{pq}}\right).$$

[Thus, $\text{Power}(.85) \doteq P\left(N(0,1) > \frac{-85 + 75 + 7.12}{3.57}\right) = P(Z > -.806) = .789$.]

Note how the power changes with n, α, p, and p_0. If you don't have enough power in a particular situation, you must increase either α or n.

This is a one parameter dist

Example 3 (Power in the normal model, σ unknown)

Consider testing $H_0: \mu \le \mu_0$ vs $H_a: \mu > \mu_0$ for iid $N(\mu, \sigma^2)$ observations.

Reject H_0 in case $\bar{X}_n > \mu_0 + t^* S_n/\sqrt{n}$ = (rejection value of the decision rule)

The power of this decision rule is given by

$$\text{Power}(\mu_0) = P(\bar{X}_n > \mu_0 + t^* S_n/\sqrt{n} \mid \mu_0 \text{ is true}) = P_{\mu_0}\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)/\sigma}{\sqrt{S_n^2/\sigma^2}} > t^*\right)$$

$$= P\left(\frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}} > t^*\right) = P(\mathcal{T}_{n-1} > t^*) = \alpha .$$

Moreover,

$$\text{Power}(\mu) = P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n} > t^*\right) = P_\mu\left(\frac{\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) + \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right)}{\sqrt{S_n^2/\sigma^2}} > t^*\right)$$

$$= P\left(\frac{Z + \delta}{\sqrt{\chi^2_{n-1}/(n-1)}} > t^*\right) = P(\mathcal{T}_{n-1}(\delta) > t^*)$$

for a non-central $\mathcal{T}$ with non-centrality parameter $\delta \equiv \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}$

= (a number we look up on the chart, with $\phi \equiv \frac{\delta}{\sqrt{2}} = \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}\frac{1}{\sqrt{2}}$).

This is a "useful" problem, but it requires charts/tables/programs — since calculations are difficult. But there is another difficulty, as δ measure the difference $\mu - \mu_0$ in units of the unknown standard deviation σ; however, this is the only solvable problem, so we make do with it

$$\mathcal{T}_m(\delta) \cong \frac{N(\delta, 1)}{\sqrt{\chi^2_m/m}} \cong \frac{Z + \delta}{\sqrt{\chi^2_m/m}}$$

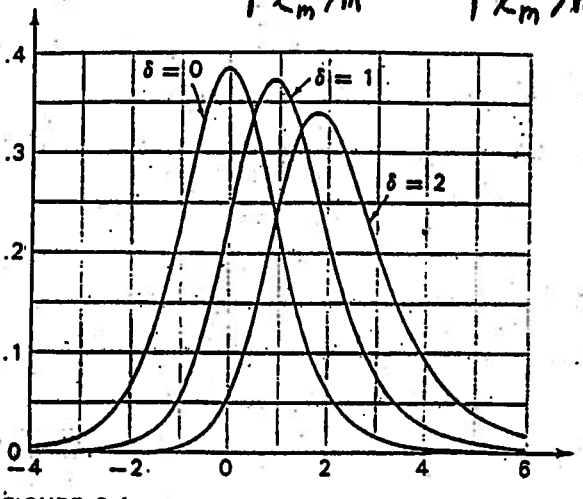
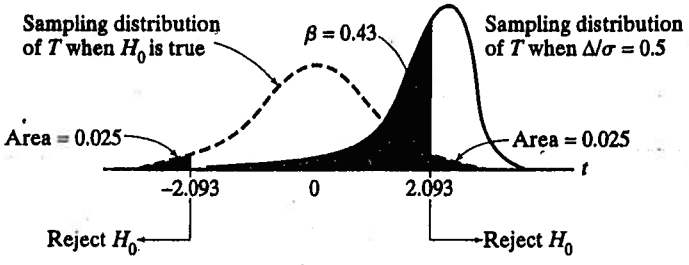


FIGURE S.4
Noncentral Student(6) distribution with noncentrality δ.



Example 4. (Power of the normal scale test)

Let $X_1, \ldots, X_n$ be independent $N(\mu, \sigma^2)$.

To test $H_0: \sigma = \sigma_0$ vs $H_a: \sigma = \sigma_1 > \sigma_0$, a natural decision rule is:

Reject $H_0: \sigma = \sigma_0$ in case $S_n^2/\sigma_0^2 = \frac{1}{n-1} SS_{XX}/\sigma^2 > \frac{1}{n-1} \chi^{2\,*}_{n-1} \equiv \frac{1}{n-1} c^*$.

The power function of this decision rule is given by

$$\text{Power}(\sigma_0) = P(\text{Reject } H_0 \mid H_0 \text{ is true}) = P\left(S_n^2/\sigma_0^2 > \tfrac{1}{n-1}\chi^{2\,*}_{n-1} \mid \sigma_0 \text{ is true}\right)$$

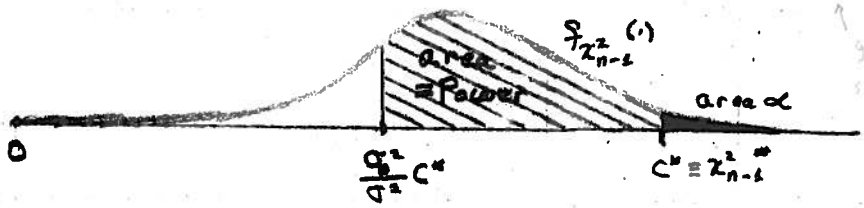
$$= P(\chi^2_{n-1} > \chi^{2\,*}_{n-1}) = \alpha \text{ , just as we set it up.}$$

and

$$\text{Power}(\sigma) = P(\text{Reject } H_0 \mid \sigma \text{ is true}) = P\left(S_n^2/\sigma_0^2 > \tfrac{1}{n-1}\chi^2_{n-1} \mid \sigma \text{ is true}\right)$$

$$= P_\sigma\left(S_n^2/\sigma^2 > \frac{\sigma_0^2}{\sigma^2}\frac{1}{n-1}\chi^2_{n-1}\right) = P\left(\tfrac{1}{n-1}\chi^2_{n-1} > \frac{\sigma_0^2}{\sigma^2}\frac{1}{n-1}\chi^2_{n-1}\right)$$

$$= P\left(\chi^2_{n-1} > \frac{\sigma_0^2}{\sigma^2}\chi^{2\,*}_{n-1}\right) \text{ or } P\left(\chi^2_{n-1} > \frac{\sigma_0^2}{\sigma^2}c^*\right).$$



(As n gets larger, the picture changes with n — unlike the 2nd picture in Example 2)

Example 5. (Power, re change of rate in a Poisson process)

A Poisson process (with intensity ν_0) runs until the time W_m of the m^{th} event. At this point something happens (the FAA changes regulations, etc.) that may have changed it to a Poisson process of intensity ν_1; and this new process is watched until the time W_n of its n^{th} failure. We wish to test $H_0: \nu_0 = \nu_1$ vs $H_1: \nu_1 < \nu_0$ (less air accidents under the new regulations).

Proposed decision rule: Reject H_0 if $\dfrac{W_m/2m}{W_n/2n} > \mathcal{F}^*_{2m,2n}$.

Show that the power of this test satisfies

$$\text{Power}(\nu_0/\nu_1) = P\left(\mathcal{F}_{2m,2n} > \frac{\nu_1}{\nu_0}\mathcal{F}^*_{2m,2n}\right).$$

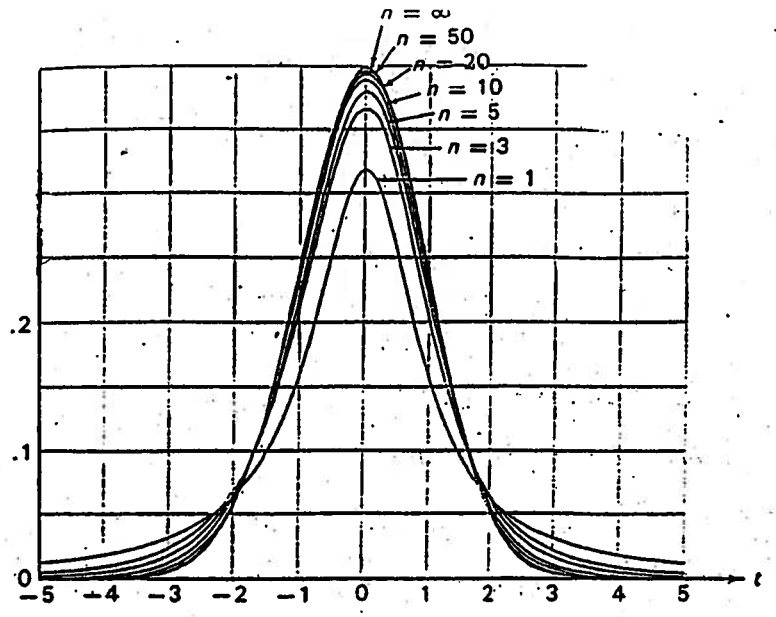


FIGURE 2.24
Student(n) distribution with $n = 1, 3, 5, 10, 20, 50, \infty$.

$$T_m(\delta) \cong \frac{N(\delta,1)}{\sqrt{\chi^2_m/m}} \cong \frac{Z+\delta}{\sqrt{\chi^2_m/m}}$$

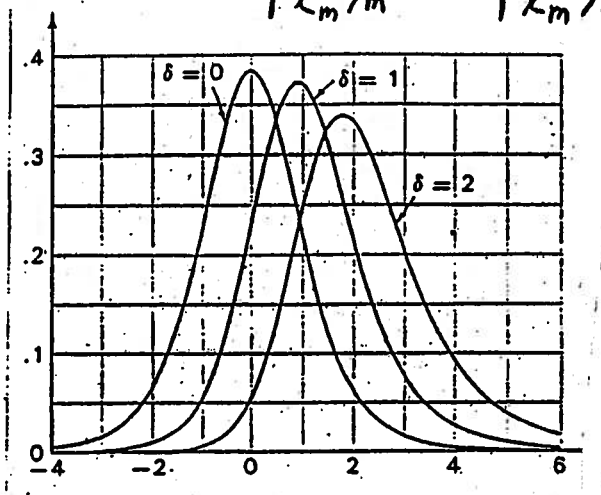


FIGURE S.4
Noncentral Student(6) distribution with noncentrality δ

$$\chi^2_m(\delta) \cong (Z_1+\delta_1)^2 + \cdots + (Z_m+\delta_m)^2$$
$$\cong (Z_1+\delta)^2 + Z_2^2 + \cdots + Z_m^2$$
where $\delta \equiv \sqrt{\delta_1^2+\cdots+\delta_m^2}$

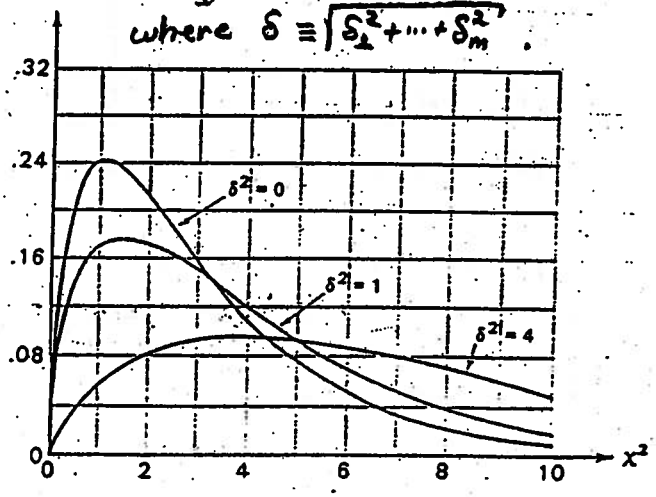


FIGURE S.5
Noncentral chi-square(3) distribution with noncentrality δ

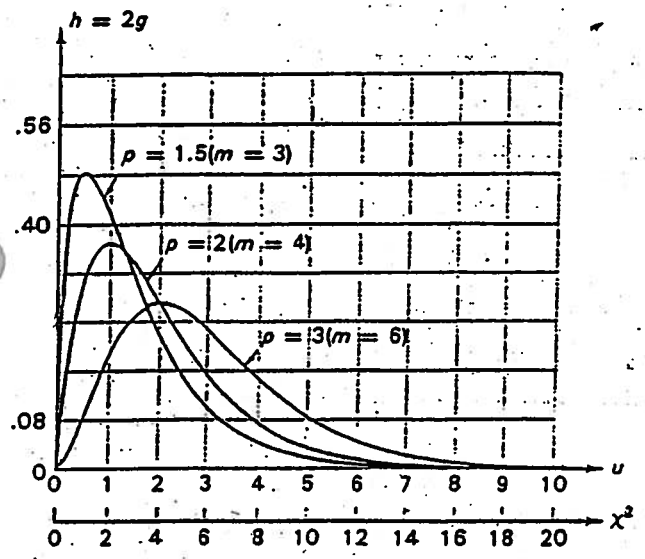


FIGURE 2.21
Gamma(ρ) distribution for u with $\rho = 1.5, 2, 3$ and chi-square(m) distributic

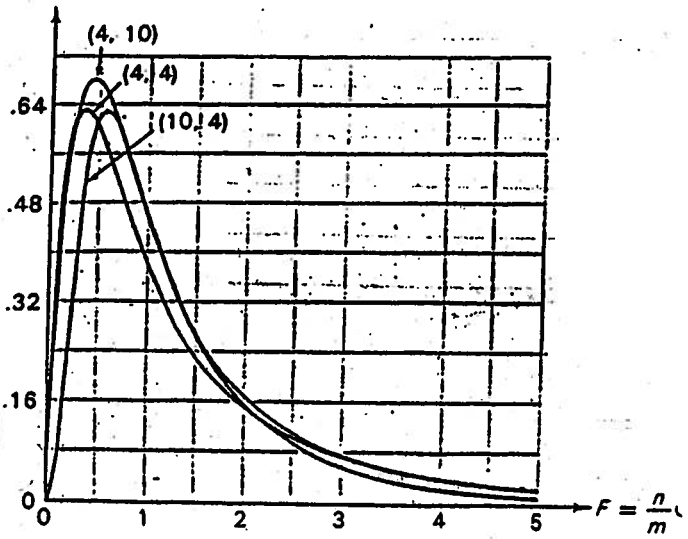


FIGURE 2.23
$F(m, n)$ distribution; $(m, n) = (4, 10), (4, 4), (10, 4)$.

$$F_{m,n}(\delta) \cong \frac{\chi^2_m(\delta)/m}{\chi^2_n/n}$$

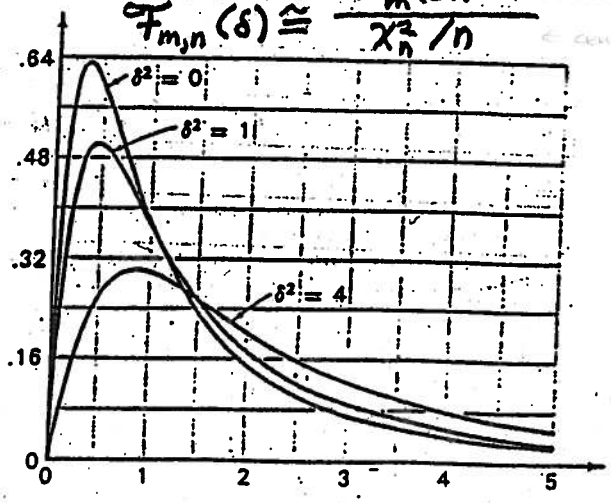


FIGURE S.6
Noncentral F(4, 4) distribution with noncentrality δ

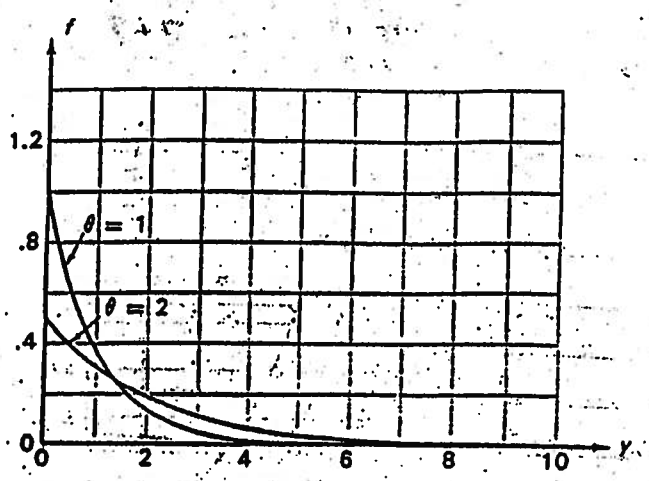


FIGURE 2.13
Exponential(θ) distribution: $\theta = 1, 2$.

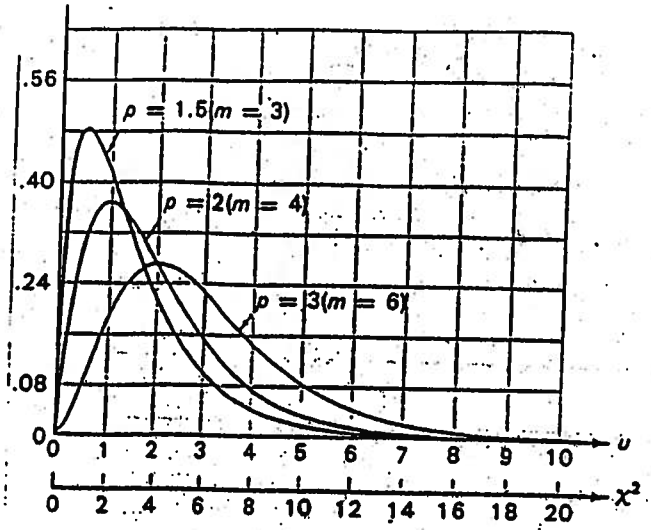


FIGURE 2.21
Gamma(ρ) distribution for u with $\rho = 1.5, 2, 3$ and chi-square(m) distribution χ^2 with $m = 3, 4, 6$.

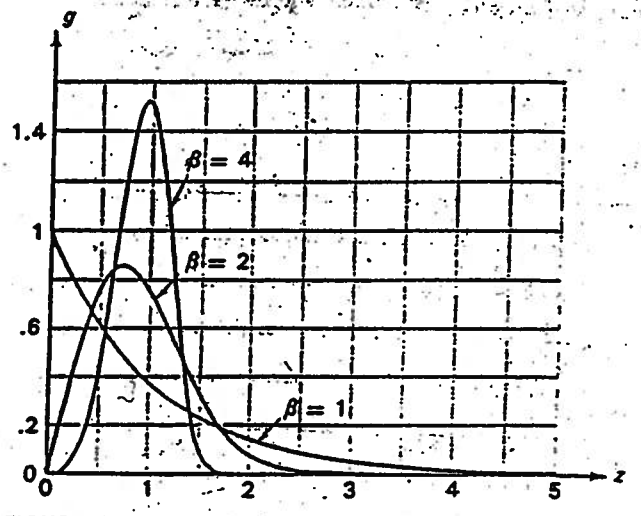


FIGURE 2.19
Weibull(β) distribution with $\beta = 1, 2, 4$.

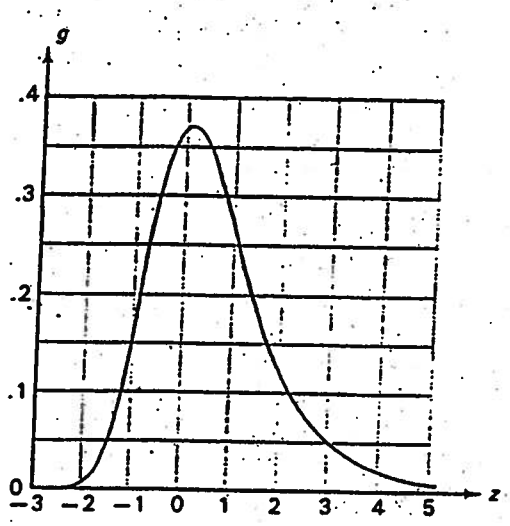


FIGURE 2.20
Standard extreme-value distribution.

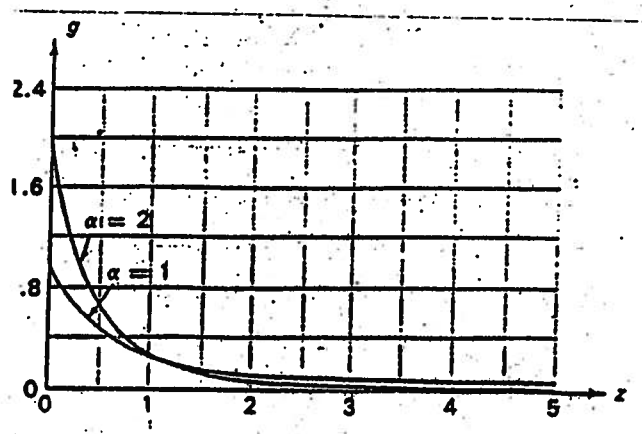


FIGURE 2.16
Pareto(α) distribution: $\alpha = 1, 2$.

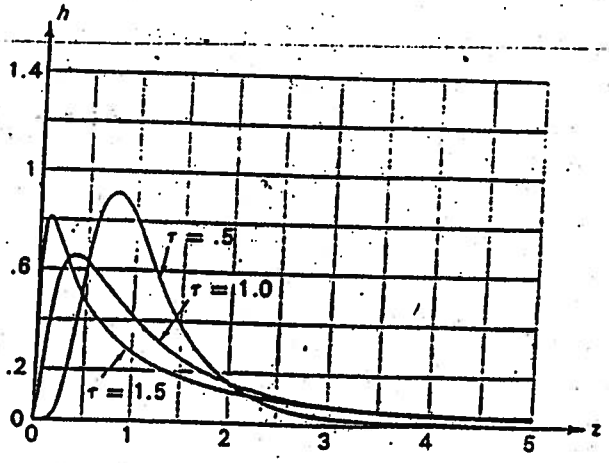


FIGURE 2.18
Lognormal(τ) distribution with $\tau = 0.5, 1.0, 1.5$.

EXAMPLE 1. (Poisson(θ)) Let $X_1, \dots, X_n$ be iid Poisson(θ). To test

(a) H_0: a fixed θ_0 is true vs H_1: a fixed θ is true, with $\theta > \theta_0$.

The Neyman-Pearson statistic is

(b) $$\frac{f_\theta(\underline{x})}{f_{\theta_0}(\underline{x})} = \frac{L(\theta; \underline{x})}{L(\theta_0; \underline{x})} = \frac{\prod_1^n e^{-\theta}\theta^{x_i}/x_i!}{\prod_1^n e^{-\theta_0}\theta_0^{x_i}/x_i!} = e^{-n\theta_0\left(\frac{\theta}{\theta_0}-1\right)}\left(\frac{\theta}{\theta_0}\right)^{\sum_1^n x_i}$$

and the NP-test rejects if this quantity is "too big". Equivalently, taking logs, the NP-test rejects if $\underbrace{-n\theta_0\left(\frac{\theta}{\theta_0}-1\right)}_{\text{constant}} + \left(\sum_1^n x_i\right)\underbrace{\log\left(\frac{\theta}{\theta_0}\right)}_{\text{positive constant}}$ is "too big." So, equivalently:

(c) The NP-test rejects if $\sum_1^n x_i > k$, where $k = n\theta_0 + z^{(\alpha)}\sqrt{n\theta_0}$ makes the test (almost) level α.

That is,

$$P_{\theta_0}(\text{Reject } H_0) = P_{\theta_0}\left(\sum_1^n x_i > n\theta_0 + z^{(\alpha)}\sqrt{n\theta_0}\right) = P_{\theta_0}\left(\frac{\sum_1^n x_i - n\theta_0}{\sqrt{n\theta_0}} > z^{(\alpha)}\right) \doteq P(N(0,1) > z^{(\alpha)}) = \alpha$$

Moreover,

$$\text{Power}(\theta) = P_\theta\left(\sum_1^n x_i > n\theta_0 + z^{(\alpha)}\sqrt{n\theta_0}\right) = P_\theta\left(\frac{\sum_1^n x_i - n\theta}{\sqrt{n\theta}} > \frac{z^{(\alpha)}\sqrt{n\theta_0} - n(\theta-\theta_0)}{\sqrt{n\theta}}\right)$$

(d) $$\doteq P\left(N(0,1) > z^{(\alpha)}\sqrt{\frac{\theta_0}{\theta}} - \frac{\sqrt{n}(\theta-\theta_0)}{\sqrt{\theta}}\right).$$

Note from (b), that the Poisson(θ) family has MLR in $\sum_1^n x_i$; thus:

(e) The NP-test of (c) is UMP level (almost) α for $H_{null}: \theta \le \theta_0$ vs $H_{alt}: \theta > \theta_0$.

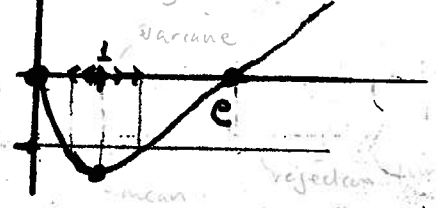
Note that the likelihood ratio LR-test of $H^0: \theta = \theta_0$ vs $H^1: \theta \neq \theta_1$ rejects H_0 when

(f) $$\lambda_n \equiv \frac{L(\hat\theta_n; \underline{x})}{L(\theta_0; \underline{x})} = e^{-n\theta_0\left(\frac{\bar x_n}{\theta_0}-1\right)}\left(\frac{\bar x_n}{\theta_0}\right)^{n\bar x_n} \text{ is "too big"};$$

or equivalently, H_0 is rejected when $\sum_1^n x_i$ is "too extreme."

Better yet, let $a \equiv n\theta_0$ and $T \equiv \bar x_n/\theta_0$; then λ_n "too big" is equivalent to

$$\frac{1}{a}\log\lambda_n = \frac{-a + a\log T}{a} = T(\log T - 1) \text{ "too big"}.$$



So the LR-test rejects $\sqrt{n}(\bar x_n - \theta_0)/\sqrt{\theta_0}$ is "too extreme."

The common compromise is to use the "equal tailed test" and reject if $\sqrt{n}(\bar x_n - \theta_0)/\sqrt{\theta_0} > z_{\alpha/2}$.

$E_{\theta_0}(T) = E_{\theta_0}(\bar x_n/\theta_0) = 1$

$\text{Var}_{\theta_0}[T] = \frac{1}{n\theta_0} \to 0$ as $n \to \infty$

$$\rightarrow t = \frac{\bar{x}_0}{\theta_0} - 1$$

$$= \frac{\bar{x} - \theta_0}{\theta_0}$$ ← subtracting mean

too big has approx $N(0,1)$ region if $\sqrt{n}\left[\frac{\bar{x} - \theta_0}{\theta_0 \sqrt{n}}\right]$

Likelihood ratio test

(H) parameter space for one model

$X_1, \ldots, X_m \sim N(\mu, \sigma^2)$

$Y_1, \ldots, Y_n \sim N(\nu, \sigma^2)$

$H_0: \mu \geq \nu \qquad H_a: \mu < \nu$

Determine a reasonable test

$$NP: \frac{\prod_1^m \frac{1}{\sqrt{2\pi}\, s_x} e^{-(x_i - \mu)^2 / 2 s_x^2}}{\prod_1^n \frac{1}{\sqrt{2\pi}\, s_y} e^{-(y_i - \bar{y})^2 / 2 s_y^2}}$$ (?)

MLE

$\hat{\mu} = \bar{x}$

$\hat{\nu} = \bar{y}$

$\hat{\sigma}^2 = \frac{SS_{xx} + S_{yy}}{m+n}$

(H) parameter space for another model

$\begin{cases} x_1, \ldots x_m \\ y_1, \ldots y_n \end{cases}$

If ν is unknown, do we have to estimate it before testing these hypotheses? Same? for σ^2?

MLE

$\hat{\mu} = \bar{z}_n = \frac{m}{N}\bar{x} + \frac{n}{N}\bar{y}$

$\hat{\sigma}^2 = \frac{SS_{zz}}{N} + \frac{SS_{zy}}{m+n}$

$= \sum_1^m (x_i - \bar{z})^2 + \sum_1^n (y_i - \bar{z})^2$

subtract $-\bar{x} + \bar{x} \qquad -\bar{y} + \bar{y}$

Want to test $\mu \geq \nu$ $\quad H_0: \mu - \nu \geq 0$

Reject if $\frac{(\bar{x} - \bar{y}) - (\mu - \nu)_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{m} + \frac{1}{n}\right)}}$ ← σ not known, replace with T dist

LIKELIHOOD RATIO TEST

MODEL. Let $X_1, \dots, X_n$ be iid with density $f_\theta(\cdot)$ for some unknown θ in the parameter space Θ. Suppose Θ has dimension r, and the subspace Θ_0 has dimension q.

(1) The LR-test of $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta \setminus \Theta_0$ is given by the decision rule:

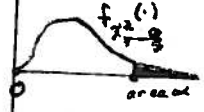
(2) Reject H_0 if $\Lambda_n \equiv \Lambda_n(\underset{\sim}{X}) \equiv \dfrac{\sup_{\theta \in \Theta} L(\theta; \underset{\sim}{X})}{\sup_{\theta \in \Theta_0} L(\theta; \underset{\sim}{X})} = \dfrac{L(\hat\theta_n; \underset{\sim}{X})}{L(\hat{\hat\theta}_n; \underset{\sim}{X})}$ is "too big"

where $\hat\theta_n$ denotes the MLE under Θ and $\hat{\hat\theta}_n$ denotes the MLE under Θ_0.

How big is "too big?" We will illustrate the following theorem.

THEOREM 1. $2\log\Lambda_n \xrightarrow{d} \chi^2_{r-q}$ as $n \to \infty$ provided that H_0 is true.

COROLLARY. The LR-test of level α rejects H_0 if $2\log\Lambda_n > \chi^2_{r-q}(\alpha)$.



[Traditionally, (old Λ) = (1/this Λ); and so $-2\log(\text{old }\Lambda) \xrightarrow{d} \chi^2_{r-q}$. New is better; it looks more like NP lemma]

EXAMPLE 1. (Geometric) Let $X_1, \dots, X_n$ be iid $\text{GeoT}(\theta)$, so $f_\theta(x) = \theta(1-\theta)^{x-1}$ for $x = 1, 2, 3, \dots$.

To test $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$. (So, $r = 1$, $q = 0$ and $r - q = 1$.) [Fact: $\bar X \doteq \left(\frac{1}{\theta}, \frac{1-\theta}{n\theta^2}\right)$]

[If the alternative had been $H_1: \theta > \theta_0$, then Theorem 1 would be false (note, $[\theta_0, \infty)$ is not a space)]

Now

(a) $L(\theta; \underset{\sim}{X}) = \prod_{i=1}^n \{\theta(1-\theta)^{X_i - 1}\} = \theta^n (1-\theta)^{\sum X_i - n}$; and so $\dot\ell(\theta; \underset{\sim}{X}) = -\frac{n}{1-\theta}\left[\bar X_n - \frac{1}{\theta}\right]$;

thus

(b) $\hat\theta_n = (\text{the MLE of } \theta \text{ under } H_0 \cup H_1) = 1/\bar X_n$, while $\hat{\hat\theta}_n = (\text{the MLE of } \theta \text{ under } H_0) = \theta_0$.

The LR-statistic is thus

(c) $$\Lambda_n = \frac{L(\hat\theta_n)}{L(\hat{\hat\theta}_n)} = \frac{L(1/\bar X_n)}{L(\theta_0)} = \frac{(1/\bar X_n)^n (1 - 1/\bar X_n)^{n(\bar X_n - 1)}}{\theta_0^n (1-\theta_0)^{n(\bar X_n - 1)}} = \frac{1}{(\theta_0 \bar X_n)^n}\left[\frac{1 - 1/\bar X_n}{1 - \theta_0}\right]^{n(\bar X_n - 1)}.$$

And so

$$2\log\Lambda_n = -2\left[n\log(\theta_0 \bar X_n) + n(\bar X_n - 1)\log\left(\frac{1-\theta_0}{1 - 1/\bar X_n}\right)\right]$$

(d) $$= -2n\left\{\log(1 + (\theta_0 \bar X_n - 1)) + (\bar X_n - 1)\log\left(1 + \left(\frac{1-\theta_0}{1 - 1/\bar X_n} - 1\right)\right)\right\} \qquad \left(\frac{1-\theta_0}{1 - 1/\bar X_n} - 1 = \frac{1 - \theta_0 \bar X_n}{\bar X_n - 1}\right)$$

$$\doteq -2n\left\{\left[(\theta_0 \bar X_n - 1) - \tfrac{1}{2}(\theta_0 \bar X_n - 1)^2 + \text{O}_{\text{ish}}\right] + (\bar X_n - 1)\left[\frac{1 - \theta_0 \bar X_n}{\bar X_n - 1} - \tfrac{1}{2}\left(\frac{1 - \theta_0 \bar X_n}{\bar X_n - 1}\right)^2 + \text{O}_{\text{ish}}\right]\right\}$$

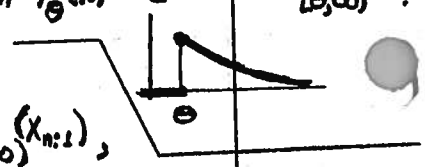
(← since $\log(1+y) \doteq y - \frac{1}{2}y^2 + \text{O}_{\text{ish}}$ when y is tiny)

$$= n(\theta_0 \bar X_n - 1)^2\left[1 + \frac{1}{\bar X_n - 1}\right] \underset{a}{=} \theta_0^2 n(\bar X_n - \tfrac{1}{\theta_0})^2\left[1 + \frac{1}{\frac{1}{\theta_0} - 1}\right]$$

(e) $$= \left\{\frac{\sqrt{n}(\bar X_n - \frac{1}{\theta_0})}{\sqrt{(1-\theta_0)/\theta_0^2}}\right\}^2 = \left\{\frac{\sqrt{n}(\bar X_n - \mu_X)}{\sigma_X}\right\}^2 \xrightarrow{\text{CLT}_d} \{N(0,1)\}^2 \overset{\text{Mann-Wald}}{=\!=} \chi^2_1 = \chi^2_{r-q}$$

when Θ_0 is true. ∎

p 376-8 **EXAMPLE 2.** (Nonstandard) Let $X_1, \ldots, X_n$ be iid with $f_\theta(x) = e^{-(x-\theta)} 1_{[\theta,\infty)}(x)$.



The likelihood function is

(a) $L(\theta; \underset{\sim}{X}) = e^{-\sum_1^n (X_i - \theta)} 1_{[\theta,\infty)}(X_{n:1}] = e^{-n(\bar{X}_n - \theta)} 1_{[\theta,\infty)}(X_{n:1})$,

and we wish to test $H_0: \theta \le \theta_0$ vs $H_1: \theta > \theta_0$. Note that

$L(\theta; \underset{\sim}{X})$ is ↑ in θ on the interval $(-\infty, X_{n:1}]$; so $\hat{\theta}_n = X_{n:1}$, and

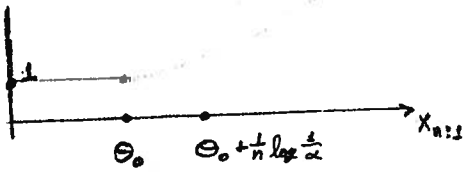
$L(\hat{\theta}_n; \underset{\sim}{X}) = e^{-n(\bar{X}_n - X_{n:1})}$.

However, under H_0,

if $X_{n:1} = \begin{cases} \le \theta_0, & \text{then } \hat{\hat{\theta}}_n = X_{n:1} \\ > \theta_0, & \text{then } \hat{\hat{\theta}}_n = \theta_0 \end{cases}$ and $L(\hat{\hat{\theta}}_n; \underset{\sim}{X}) = \begin{cases} e^{-n(\bar{X}_n - X_{n:1})} \\ e^{-n(\bar{X}_n - \theta_0)} \end{cases}$.

Thus

(b) $\Lambda_n = \begin{cases} 1 & \text{if } X_{n:1} \le \theta_0 \\ e^{n(X_{n:1} - \theta_0)} & \text{if } X_{n:1} \ge \theta_0 \end{cases}$.



Well, $P_{\theta_0}(\Lambda_n \ge \frac{1}{\alpha}) = P_{\theta_0}(e^{n(X_{n:1} - \theta_0)} \ge e^{\log 1/\alpha}) = P_{\theta_0}(X_{n:1} \ge \theta_0 + \frac{1}{n}\log\frac{1}{\alpha})$

$= P(\text{Min of n rv's dist as Exp}(1) \ge \frac{1}{n}\log\frac{1}{\alpha})$

p. 386 (c) $= P(\text{Exp}(1) \ge \log\frac{1}{\alpha}) = \exp(-\log\frac{1}{\alpha}) = \alpha = (\text{the level of the test})$.

Thus the decision rule that rejects H_0 when $\lambda_n \ge \frac{1}{\alpha}$ (or better, when $X_{n:1} \ge \theta_0 + \frac{1}{n}\log\frac{1}{\alpha}$) is level α; and

$\text{Power}(\theta) = P_\theta(e^{n(X_{n:1} - \theta) + n(\theta - \theta_0)} \ge e^{\log 1/\alpha}) = P(\text{Exp}(1) \ge (\log\frac{1}{\alpha}) - n(\theta - \theta_0))$

(d) $= \begin{cases} \alpha e^{n(\theta - \theta_0)} & \text{if } \theta_0 \le \theta \le \theta_0 + \frac{1}{n}\log\frac{1}{\alpha} \\ 1 & \text{if } \theta > \theta_0 + \frac{1}{n}\log\frac{1}{\alpha} \end{cases}$.

Comment: what is the asymptotic distribution of $\hat{\theta}_n$ in this nonstandard case?

Well,

$P_\theta(n(\hat{\theta}_n - \theta) > t) = P_\theta(\hat{\theta}_n > \theta + \frac{t}{n}) = P_\theta(X_{n:1} - \theta > \frac{t}{n}) = P(E_{n:1} > \frac{t}{n})$

$= P(\text{Exp}(1) > t/n)^n = (e^{-t/n})^n = e^{-t} = P(\text{Exp}(1) > t)$.

That is,

(e) $n(\hat{\theta}_n - \theta) \to_d \text{Exp}(1)$ as $n \to \infty$ (when θ is true).

p.376 **EXAMPLE 3.** (Normal theory) Let $X_1,\dots,X_n$ be iid $N(\theta,\sigma_0^2)$; so $r=1$ and $q=0$.

The LR-test for $H_0:\theta=\theta_0$ vs $H_1:\theta\neq\theta_0$ is given by:

(2) Reject H_0 if $\sqrt{n}\,|\bar{X}_n-\theta_0|/\sigma_0 \geq z^{(\alpha/2)}$.

Note that $\hat{\theta}_n=\bar{X}_n$ (as shown in MLE section), and so the LR-test rejects H_0 if

$$\Lambda_n=\frac{(2\pi\sigma_0^2)^{-n/2}\exp\left(-\frac{1}{2\sigma_0^2}\sum_1^n(X_i-\bar{X})^2\right)}{(2\pi\sigma_0^2)^{-n/2}\exp\left(-\frac{1}{2\sigma_0^2}\sum_1^n(X_i-\theta_0)^2\right)}=\exp\left(\frac{1}{2\sigma_0^2}\left\{\sum_1^n(X_i-\theta_0)^2-\sum_1^n(X_i-\bar{X})^2\right\}\right)$$

$$=\exp\left(\frac{1}{2\sigma_0^2}n(\bar{X}-\theta_0)^2\right) \quad \text{is "too big"},$$

as claimed. Note,

$$2\log\Lambda_n=\left[\frac{\sqrt{n}(\bar{X}_n-\theta_0)}{\sigma_0}\right]^2\cong Z^2\cong\chi^2_1 \text{ under } H_0. \quad (\text{Here } r_0\equiv r \text{ and } g_0\equiv g.$$

EXAMPLE 4. (Normal theory) Let $X_1,\dots,X_n$ be iid $N(\theta,\sigma^2)$.

The LR-test of $H_0:\theta=\theta_0$ vs $H_1:\theta\neq\theta_0$ (with $r=2$, $g=1$, and $r-g=1$) is given by:

(3) Reject H_0 if $\sqrt{n}\,|\bar{X}_n-\theta_0|/S_n \geq t^{(\alpha/2)}_{n-1}$.

Note that: $\hat{\theta}_n=\bar{X}_n$ and $\hat{\sigma}_n^2=\frac{1}{n}\sum_1^n(X_i-\bar{X})^2$, while $\hat{\hat{\theta}}_n=\theta_0$ and $\hat{\hat{\sigma}}_n^2=\frac{1}{n}\sum_1^n(X_i-\theta_0)^2$.

Thus the LR-test rejects H_0 if

$$\Lambda_n=\frac{(2\pi\hat{\sigma}_n^2)^{-n/2}\exp\left(-\frac{1}{2\hat{\sigma}_n^2}\sum_1^n(X_i-\bar{X})^2\right)}{(2\pi\hat{\hat{\sigma}}_n^2)^{-n/2}\exp\left(-\frac{1}{2\hat{\hat{\sigma}}_n^2}\sum_1^n(X_i-\theta_0)^2\right)}=\left[\frac{\hat{\sigma}_n^2}{\hat{\hat{\sigma}}_n^2}\right]^{\frac{n}{2}}\frac{e^{-n/2}}{e^{-n/2}}=\left[\frac{\hat{\sigma}_n^2}{\hat{\hat{\sigma}}_n^2}\right]^{\frac{n}{2}} \text{ is "too big".}$$

Equivalently,

Reject H_0 if $(n-1)\left(\Lambda_n^{\frac{2}{n}}-1\right)=(n-1)\dfrac{\hat{\hat{\sigma}}_n^2-\hat{\sigma}_n^2}{\hat{\sigma}_n^2}=\dfrac{\sum_1^n(X_i-\theta_0)^2-\sum_1^n(X_i-\bar{X})^2}{\frac{1}{n-1}\sum_1^n(X_i-\bar{X})^2}=\dfrac{n(\bar{X}-\theta_0)^2}{S_n^2}=T_n^2$

is "too big"

or if $\sqrt{n}\,|\bar{X}_n-\theta_0|/S_n > t^{(\alpha/2)}_{n-1}$.

Note that

$$2\log\Lambda_n=2\cdot\frac{n}{2}\log\left(1+\frac{\hat{\hat{\sigma}}_n^2-\hat{\sigma}_n^2}{\hat{\sigma}_n^2}\right)\underset{n}{=}(n-1)\frac{(\hat{\hat{\sigma}}_n^2-\hat{\sigma}_n^2)}{\hat{\sigma}_n^2}=\frac{\sum_1^n(X_i-\theta_0)^2-\sum_1^n(X_i-\bar{X})^2}{S_n^2}=\frac{n(\bar{X}-\theta_0)^2}{S_n^2}=T_n^2$$

Thus $2\log\Lambda_n\underset{n}{=}(n-1)\left(\Lambda_n^{2/n}-1\right)=\left\{\dfrac{\sqrt{n}(\bar{X}-\theta_0)}{S_n}\right\}^2\to_d\{N(0,1)\}^2=\chi^2_1=\chi^2_{r-g}=\chi^2_{r_0-g_0}$

Also, $\dfrac{(n-1)}{1}\left(\Lambda_n^{\frac{2}{n}}-1\right)=\left\{\dfrac{\sqrt{n}(\bar{X}-\theta_0)}{S_n}\right\}^2\cong\{\text{Student's } t_{n-1}\}^2\cong(\text{Snedecor's } F_{1,n-1})$

These last facts have strong parallels all through ANOVA; that is,

(4) $\dfrac{(n-r_0)}{(r_0-g_0)}\left(\Lambda^{\frac{2}{n}}-1\right)\cong(\text{Snedecor's } F_{r_0-g_0,\,n-r_0})$ for $r_0\equiv r-1$ and $g_0\equiv g-1$

2-6-2004

Example

2-6-2004

Test, best possible test, try to get a lower bound

reject if too big

Needs to go to $\bar{X}$ or something familiar

How to get rich? Realestate

$x \in S$ ← sample space

↑ houses

α - my budget

$f_{\theta_0}(x) = ($ the cost of a specific house $x)$

Subject to side condition: $\sum_{x \in C} f_0(x) \leq \alpha$ ← can buy this many houses

where C is the critical set of houses one buys

- Buy which houses? Want maximize profits

$f_\theta(x) = ($ the ~~value~~ projected value at 5 years$)$

which houses does one buy, put in my critical region?

- Buy houses for which $\lambda = \frac{f_\theta(x)}{f_{\theta_0}(x)} > C$ is big

best possible thing one can do

f_{θ_0}, f_θ is an assume distribution

UMP - uniformly most powerful

maximized likelyhood of all possible θ also known as $\bar{X}_n$

3-6-2004 HW

#19 $Y_i = \beta X_i + \epsilon_i$ for ϵ_i's iid $N(0, \sigma^2)$

$N(\beta X_i, \sigma^2)$

Sufficient & complete

$$f_Y(y) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_1^n (Y_i - \beta X_i)}$$

$$= (\quad) e^{\left[-\frac{1}{2\sigma^2}\right]\left[\sum X_i^2\right] + \left[\frac{\beta}{\sigma^2}\right]\left[\sum X_i Y_i\right]\left(e^{-\frac{1}{2\sigma^2}\beta^2 \sum_1^n X_i^2}\right)}$$

n_2 T_2 n_1 T_1

will contain a rectangle sufficient & complete

statistic based on a sufficient statistic is still sufficient

$$0 = \dot{l}_\beta = \frac{\partial}{\partial \beta} \log L = \frac{\partial}{\partial \beta} \log f$$

$$0 = \dot{l}_{\sigma^2}$$

$$\Sigma (y_i - \beta x_i) x_i = 0$$

Can only measure true with error

truth is parameter θ in statistic

model eqn

$$y_i = \theta + \epsilon_i \quad \sim N(\theta, \sigma^2)$$

one simple normal problem

$\bar{X}$, s^2

mean θ σ^2

var $\frac{\sigma^2}{n}$ hard

$$T(\theta) = \pi \theta^2$$

$$T_{guess} = \pi \bar{X}^2 - \frac{\pi s^2}{n}$$

$$E\hat{T} = \pi \left(\theta^2 + \frac{\sigma^2}{n}\right)$$

to fix

$$P(\min \text{ of } \exp_i > t) = P(\text{all } \exp_i > t) = P^n(E_i > t)$$

$$= (e^{-t})^n = e^{-nt}$$

STAT 506

3-6-2004

Help Session 12:30 on Tuesday.
Final 8:30 on Wednesday.

Final Exam

#1. MME, MLE, Info = $\mathcal{I}(\theta)$ and Asym Normal and confirm, CR-bnd.
(No non-standard models will be used.)

#2. A normal problem where you can begin with $\bar{X}$, S_x^2, etc, as sufficient (complete via an exp family - the seperator) and go from there.
Old + Practice + HW: Student T-technology.
(Too fast to do if long.)

#3 Simple Bayes estimator $E(\tau(\theta) \mid \underset{\sim}{x})$ ← or y = sufficient

#4. Simple UMVUE (maybe in the context of #2).
(Not via $V = E(W|T)$).

#5. LR-test via $2\log\lambda \to_d \chi^2_{r-q}$ (and confirm via CLT, maybe).

#6. One blatantly prob question (at most), post stat questions.

Ch 3.5 for probability

AMATH 506 : Shoracek Galen

Office hours 3.3-4.3 Wednesday B323 Padelford

HW : 2, 5, 11

+ means union is disjoint

$P(A+B) = P(A \cap B)$

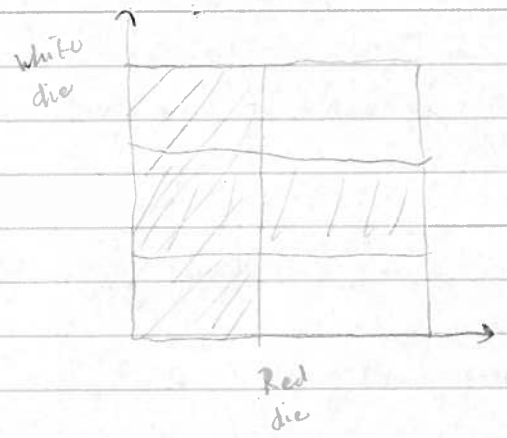
- Conditional Probability: Prob of A given B $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  Allows one to reconsualize sample space
- Independence: $P(A|B) = P(A)$
- N_{AB} = number of ways AB can happen

$$P(A/B) = \frac{P(A\,B)}{P(B)} = \frac{N_{AB}/N_B}{N_B/N_B} = \frac{N_{AB}}{N_B}$$

Divides into two trivial problems (conditional probability)

Craps win with 1st roll is 10 $P(10)$ is $\frac{3}{36}$

$P(+ \leq 3 \mid + \neq 7) = \frac{3}{30}$



Disjoint events never independent

Indep, doesn't depend on previous event

- If several experiments have same distribution, can't tell which experiment was done

- MN Rule

C_k^n - number of combinations of k things from n
n chose k

P_k - permute

$P_k^n = C_k^n k!$

HW: (Ch1) 2, 5, 11, 13, 14, 25, 33, 38 L1
17, 18, 21, 23, 24

(Ch1) 44, 83 L2
Ch2 4, 8c, 17, 19, 24

- Statistics is probability done backwards

Average value (expectation)

$E(x) = \sum_{\text{all } x} x \times P(x)$ possibility * probability

$\sum_{\text{all } k} k P(x)$

Physical interpretation → balance point
μ = mean, average

$\mu \equiv E(x) = \sum x P(x)$

Measure how far from average, how spread out data is

$\tau \equiv E(|x - \mu|)$ Not used (mean deviation)

$f(\cdot)$ — function

Any experiment looks the same

Average distance squared

$$\sigma = \sqrt{E(x-\mu)^2}$$

$\sigma_1^2 + \dots + \sigma_2^2$ — variance of the sum is the sum of the variances IF INDEP!

Std dev. important for meaning

$$E(g(x)) = \int_{\text{adm.}} g(x)\, f(x)\, dx$$

$f(x)$: probability

$$E(x^2 - 2x\mu - \mu^2) = Ex^2 - E(2x\mu) + \mu^2$$

$$-2E(x)\cdot\mu$$

$$= Ex^2 - \mu^2$$

$$\text{Exp}(\gamma) = \theta = \frac{1}{\nu}$$

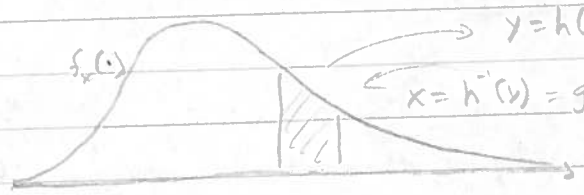
Exp(mean $= \theta$) ← handy for stats

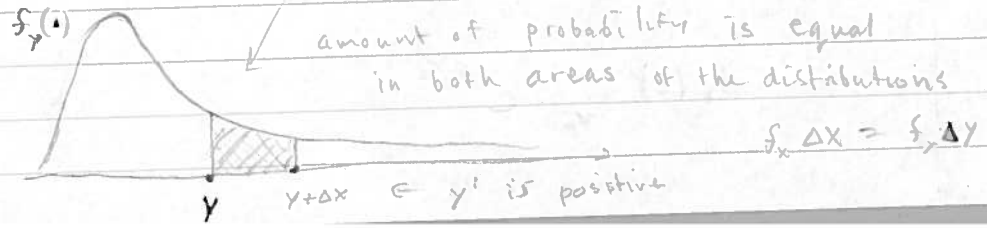
Exp(rate $= \nu$) ← handy for prob

- Probabilities are areas

- Ideas for variable changes

Random var. X, continuous



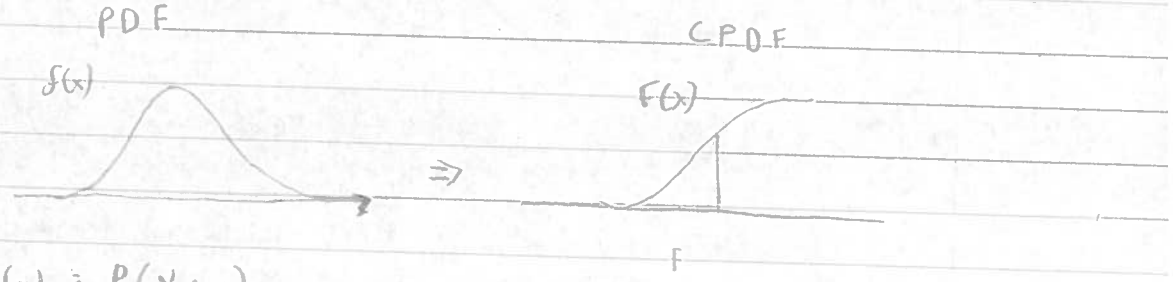


For more than dimension

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \vdots & & \\ \frac{\partial x_n}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

$|J|$ = absolute value of jacobian

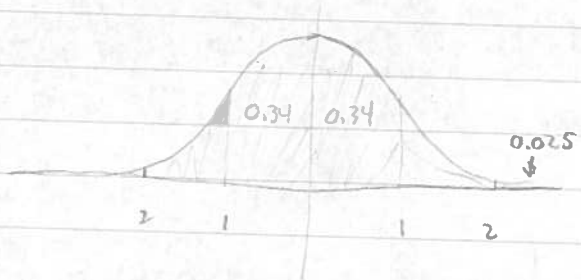
$$f(x_1, x_2, \ldots x_n) = f(y_1, \ldots, y_n)|J|$$



$$\begin{aligned} F_y(y) &= P(Y \le y) \\ &= P(h(x) \le y) \\ &= P(X \le h^{-1}(y) = g(x)) \\ &= F_x(g(x)) \end{aligned}$$

Error replace $\frac{d\theta}{dx}$ with $\frac{dx}{d\theta}$

Normal Distribution



Adjust scales by a, divide height by a

Z for normal rand. variable

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Assume 16 = R, 36 = W and N = 52

C_{13}^{52}

If want k red ones Need to apply ma rule

$C_k^{16} \cdot C_{13-k}^{36}$ possible cases

must choose total of 13 cards

$$P = \frac{C_k^R C_{n-k}^W}{C_n^N}$$

Variance $EX^2 - \mu^2$

$EX^2 = 1^2 p + 0^2 (1-p)$ one draw

$= p$

$\text{Var}[X] = EX^2 - \mu^2 = p - p^2 = p(1-p) = pq$

If find variance of one draw

n indep draws (with replacement) $\text{Var}[X] = npq$

Mean = np (always)

Variance = $npq\left[1 - \frac{n-1}{N-1}\right]$

adjusts for dependence แก้ไขสำหรับการขึ้นต่อกัน

• For indep events

N not known can't use binomial dist

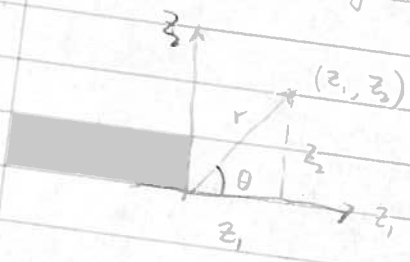
$P(k) = P(T=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k \geq 0$

Taylor series

$P_n(k) = C_k^n p^k q^{n-k}$

$q = 1 - p$

want to write joint density functions



1-4-2004

Hypergeometric

R - reds

W - whites

$N = R + W$

$X_i = \begin{cases} 1 & \text{if } R \\ 0 & \text{if } W \end{cases}$

Prob = $R/N \Rightarrow P = \frac{R}{N}$ ① ② ③ ... ⑤ balls

Sample n times

Method I : w/o ← dep

Method II : with ← indep rand variable

Marginal dist: same marginal (look at one at time)

$E(X) = \sum_{\text{all } x} x P(x) = 1 \cdot \frac{R}{N} + 1 \cdot \frac{W}{N} = 1 \cdot p + 1 \cdot q$

n turns

$E(T_n) = E(X_1 + \dots + X_n) = E(X_1) + E(X_2) + \dots E(X_n)$

$= nm = np$

Dist $\rightarrow T_n = X_1 + \dots + X_n$

$E T_n = \sum_{\text{all } t} t P(T = t)$

$P(T = k) = \frac{N_{[T=k]}}{N_S}$ ← desired result? counts number of red balls

← sample space

Ch 3.5 for probability

AMATH 506 : Shorack Galen

Office hours 3.3-4.3 Wednesday B623 Padelford

HW : 2, 5, 11

+ means union is disjoint

$P(A+B) = P(A \cap B)$

- Conditional Probability: Prob of A given B $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  Allows one to revisualize sample space
- Independence $P(A|B) = P(A)$

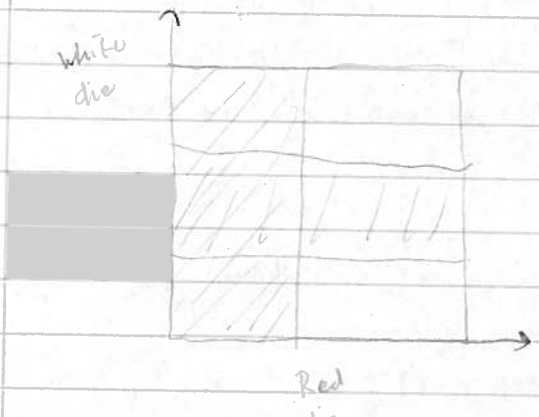
- N_{AB} = number of ways AB can happen

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{N_{AB}/N_B}{N_B/N_B} = \frac{N_{AB}}{N_B}$$

Divides into two trivial problems (conditional probability)

Craps win with 1st roll is 10 $P(10)$ is $\frac{3}{36}$

$P(\text{1st} \leq 3 \mid \text{1st} \neq 7) = \frac{3}{30}$



Disjoint events never independent

Indep, doesn't depend on previous event

- If several experiments have same distribution, can't tell which experiment was done

- MN Rule

C_k^n - number of combinations of k things from n
n chose k

P_k^n - permute

$P_k^n = C_k^n \, k!$

HW: (Ch1) 2, 5, 11, 13, 14, 25, 33, 38 — L1
17, 18, 21, 23, 24

(Ch1) 44, 53 — L2
Ch2 4, 8c, 17, 19, 24

- Statistics is probability done backwards

Average value (expectation)

$E(x) = \sum_{\text{all } x} x \, P(x)$ possibility * probability

$\sum_{\text{all } k} k \, P(x)$

Physical interpretation → balance point
μ = mean, average

$\mu \equiv E(x) = \sum x \, P(x)$

Measure how far from average, how spread out data is

$\tau \equiv E(|x - \mu|)$ Not used (mean deviation)

$f(\cdot)$ — function

Any experiment looks the same

Average distance squared

$$\sigma = \sqrt{E(x-\mu)^2}$$

$\sigma_1^2 + \dots + \sigma_2^2$ — variance of the sum is the sum of the variances IF INDEP!

Std dev. important for meaning

$$E(g(x)) = \int_{\text{adv}} g(x) f(x)\,dx$$

$f(x)$ ← probability

$$E(x^2 - 2x\mu - \mu^2) = Ex^2 - E(2x\mu) + \mu^2$$
$$-2E(x)\cdot\mu$$
$$= Ex^2 - \mu^2$$

$$\text{Exp}(\gamma) = \theta = \frac{1}{\gamma}$$

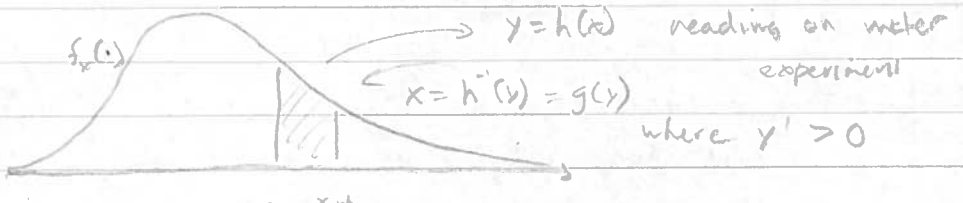
Exp(mean $= \theta$) ← handy for stats

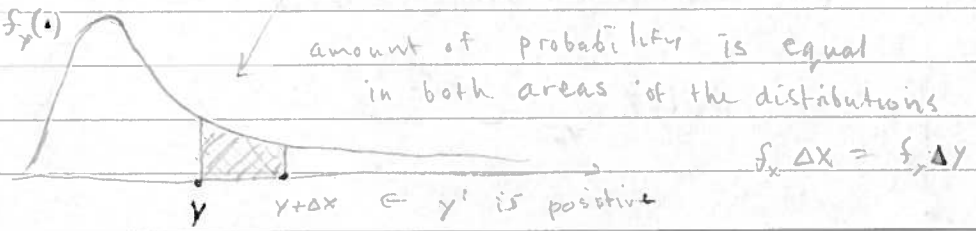
Exp(rate $= \nu$) ← handy for prob

- Probabilities are areas

- Ideas for variable changes

Random var. X, continuous



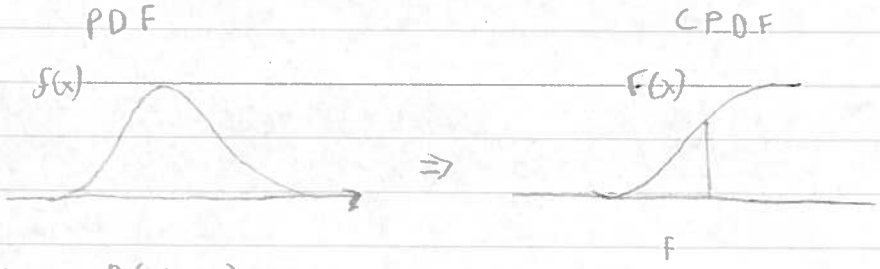


For more than dimension

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \vdots & & \\ \frac{\partial x_1}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

↙ absolute value of jacobian

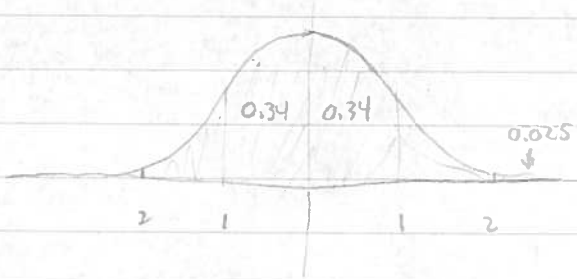
$$f(x_1, x_2, \ldots x_n) = f(y_1, \ldots, y_n)|J|$$



$$F_{y=}(y) = P(Y \le y)$$
$$= P(h(x) \le y)$$
$$= P(X \le h^{-1}(y) = g(x))$$
$$= F_x(g(y))$$

Error replace $\frac{d\theta}{dx}$ with $\frac{dx}{d\theta}$

Normal Distribution



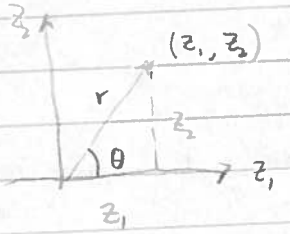
Adjust scales by a, divide height by a

Z for normal rand. variable

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$q = 1 - p$ book notation

want to write joint density functions



1-4-2004

Hypergeometric

R - reds

W - whites $\quad N = R + W$

$X_i = \begin{cases} 1 & \text{if } R \\ 0 & \text{if } W \end{cases}$

Prob = $R/N \Rightarrow P = \frac{R}{N}$ ① ② ③ ... ⑯ balls

Sample n times

Method I : w/o ← dep

Method II : with ← indep rand variable

Marginal dist: same marginal (look at one at time)

$$E(X) = \sum_{\text{all } x} x P(x) = 1 \cdot \frac{R}{N} + 1 \cdot \frac{W}{N} = 1 \cdot p + 1 \cdot q$$

n turns

$$E(T_n) = E(X_1 + \dots + X_n) = E(X_1) + E(X_2) + \dots E(X_n)$$

$$= n\mu = np$$

Dist $\rightarrow T_n = X_1 + \dots + X_n$

$$E T_n = \sum_{\text{all } t} t \, P(T = t)$$

$P(T=k) = \frac{N[T=k]}{N_s}$ ← desired result ← sample space

counts number of red balls

Assume 16 = R, 36 = W and N = 52

C_{13}^{52}

If want k red ones Need to apply mn rule

$C_k^{16} \cdot C_{13-k}^{36}$ possible cases

↑ must choose total of 13 cards ↗

$$P = \frac{C_k^R C_{n-k}^W}{C_n^N}$$

Variance $EX^2 - \mu^2$

$EX^2 = 1^2 p + 0^2 (1-p)$ one draw

$= p$

$$\text{Var}[X] = EX^2 - \mu^2 = p - p^2 = p(1-p) = pq$$

If find variance of one draw

n indep draws (with replacement) $\text{Var}[X] = npq$

Mean = np (always)

Variance = $npq\left[1 - \frac{n-1}{N-1}\right]$

adjusts for dependence แต่ไม่รู้เอาไปใช้ตอนไหน

• For indep events

N not known can't use binomial dist

$$P(k) = P(T=k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k \geq 0$$

Taylor series

$P_n(k) = C_k^n p^k q^{n-k}$.

$$P_n(\ ) = \frac{1}{k!} \frac{n(n-1)\cdots(n-k+1)}{n(n)\cdots n} (P_n n)^k \left(1 - \frac{nP_n}{n}\right)^n \frac{1}{(1-P_n)^k}$$

$\uparrow k$ ns

$$= \frac{1}{k!}(1)\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(\lambda_n)^k \left(1-\frac{\lambda_n}{n}\right)^n \frac{1}{(1-P_n)^k}$$

as $n \to \infty$

$$= \frac{1}{k!}(1^k)\lambda^k e^{-\lambda} \frac{1}{1^k}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!}$$

Mean & variance

$$\mu = \sum_{\text{all } \mu} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda$$

{ binomial $\mu = np$
poisson $np_n \to \lambda$

Variance $= \lambda$

binomial $n \cdot p_n(1-p_n) \to \lambda(1)$
$= \lambda$

Geometric (Number of times needed until success)

0 0 0 0 0 1 0 0

Mean in one turn:

ave: $\frac{1}{p} = \frac{?}{1}$

one over rate = mean

p is rate

$$P(Y=k) = q^{k-1}p$$

Sigma fields

Probability integral transformation

Let F be a d.f

Let $U \approx U_{(0,1)}$

$X = F^{-1}(U) \cong F$

Proof

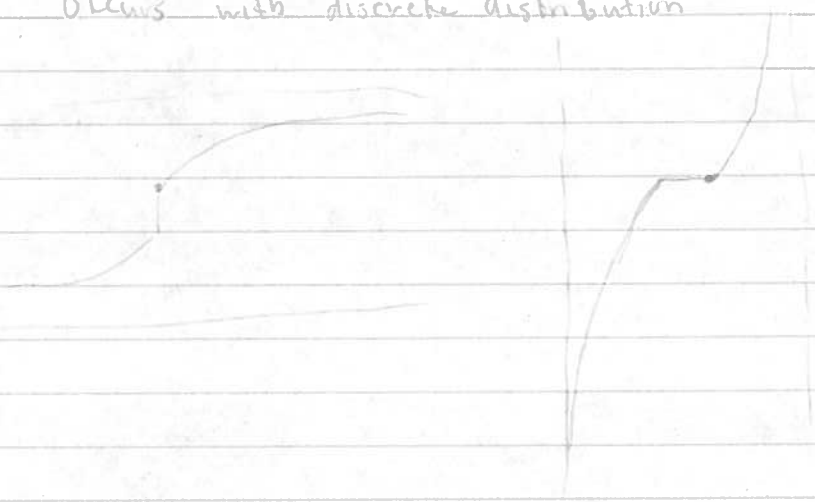
$F_X(x) = P(X \le x) = P(F^{-1}(u) \le x)$

$= P(u \le F(x)) = F(x)$

Only good for left continuous inverse.

Right

Occurs with discrete distribution



6-4-2004

Intensity p for success (intensity for success in each term)

T_n - number of successes in n turns

X_n - Bernoulli Random variable

Y - time until first success (interarrival time)

time between successes

$P(Y = k) = q^{k-1}p$ for $k = 1, 2, \ldots$

$\cong$ geometric^{r} dist (P) (r turns) pg 95 Neg Binomial

counts number of turns (until) first success

$P(Y > k) = \sum_{i=k+1}^{\infty} P(Y = i)$

$= pq^{(k+1)-1} + pq^{(k+2)-1} + \cdots$

$= pq^{k+1}$

$= \frac{pq^k}{1-q} = q^k$

redundant

$$P(Y > i+k \mid Y > i) = \frac{P(Y > i+k \cap Y > i)}{P(Y > i)}$$

$$= \frac{q^{k+i}}{q^i} = q^k$$

$$= P(Y > k)$$

- Only discrete dist that has lack of memory

- How long to wait for rth success $= W_r$

$W_r = \text{Neg Bi Turns}(r, p)$

const intensity p

$k-1$ successes before | k successes

$$P(W_r = k) = P(T_{k-1} = r-1) \text{ and } P(X_k = 1)$$

$$= P(T_{k-1} = r-1) \cdot P(X_k = 1) \quad \text{independence}$$

$$= C_{r-1}^{k-1} p^{r-1} q^{(k-1)-(r-1)} \cdot p$$

$$= \left(C_{r-1}^{k-1}\right) p^r q^{k-r} \quad \text{for } k \geq r$$

Same p and indep – ~~Binomial~~ sum of binomial is equal to binomial for sum of turns

Sum of neg binomial is neg binomial

$$\binom{7}{1} p^1 q^{7-1} \cdot p$$

138.5 Continuity correction

boxes of district

$$\frac{P\left(Y_1 + \cdots Y_{25} - rM \leq 138 - 25\frac{1}{p}\right)}{\sqrt{r\frac{q}{p^2}}} \qquad 25\frac{0.8}{(0.2)^2}$$

$$P(T_{138} > 25) = P\left(\frac{X_1 + \cdots + X_{138} - np}{\sqrt{npq}} > \frac{25}{\sqrt{138\frac{1}{5}\frac{4}{5}}}\right)$$

$\frac{1}{\sqrt{n}}$ skewness } bad for central limit theorem
$\frac{1}{n}$ heavy tails }

Geometric is very skewed
Binomial is less skewed

7-4-2004

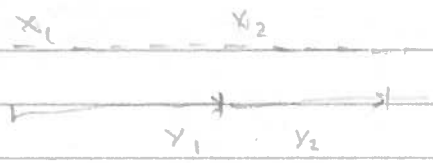
$\left(\frac{3}{7}\right)^{12}$ for no calls on two days of week

$\binom{7}{2}$ ways to select two days

$$P\left(\bigcup_1^n A\right) = \sum_1^n P(A_i) - \sum_{i<j} \sum P(A_i \cap A_j) +$$

exclusion / inclusion

T_n - num turns



inter arrival time

- waiting time for n sucesses

m sucesses where are they
each are equally likely

m - red
n-m - green

$$P(T_k = i \mid T_n = m) = \frac{C_i^m \, C_{k-i}^{n-m}}{C_k^n}$$ hypergeo.

Poisson process

q unknown, n unknown, p_n - is known

ν = intensity
= mean # of counts (events) per unit of time

N_t (# successes by time t)

$N_t \sim$ Poisson $(\lambda = \nu t)$ amount expected at time t

event $[Y_1 > t] = [N_t = 0]$

$1 - F_Y(t) = P(Y_1 > t) = P[N_t = 0] = P(\text{poisson}(\nu t) = 0)$

$= \frac{(\nu t)^0}{0!} e^{-\nu t}$

$\Rightarrow f_Y(t) = \nu e^{-\nu t}$

time increments are indep, and stationary

mean $= E_Y = \theta = \frac{1}{\nu}$

$f_Y(t) = \nu e^{-\nu t}$ for $t \geq 0$

$= \frac{1}{\theta} e^{-t/\theta}$ for $t \geq 0$

$[W_k > t] = [N_t < k] \Rightarrow P[W_k > t] = P[N_t < k]$

$= \sum_{i=0}^{k-1} P(\text{Poisson}(\nu t) = i)$

$= \sum_{i=0}^{k-1} \frac{e^{-\nu t}(\nu t)^i}{i!}$

derivative of this gives gamma dist.
wait for k hits

Poisson is only continuous distribution that has lack of memory

$P(y > t+s \mid y > s) = \frac{P(y > t+s \text{ and } y > s)}{P(y > s)}$ redundant

$= \frac{e^{-\nu(s+t)}}{e^{-\nu s}} = e^{-\nu t} = P(y > t)$

4:30 Prob session
RM C301 Padd

If n gets large, poisson looks normal

Add gammas together starts to look normal

8-4-2004

N_t, $t \geq 0$ is a Poisson process of intensity ν

Then $N_t \cong$ Poisson$(\lambda = \nu t)$ (mean amount in time interval t)

with indep. & stationary increments

Interarrival times $Y_1, Y_2, \ldots$ and Exp [ratio $= \nu$]

pdf $= f_y(y) = \nu e^{-\nu y}$ for $y \geq 0$

Waiting times $W_k = Y_1 + \ldots Y_k$

has Gamma $(k$, rate $\nu)$ $\quad f(w) = \frac{\nu^k w^{k-1} e^{\nu w}}{\Gamma(k)}$ for $w \geq 0$

Key fact $[W_t > t] = [N_t < k]$

$[W_t \geq t] \neq [N_t < k]$

$\Rightarrow$

Conditionally uniform

uniformly distributed

ordered statistics (value) put in same order, do not have to arrive in this order

Central limit theorem $\Rightarrow$ applies to distribution with μ and σ^2

Calculating β, Γ moments

$$\int_0^1 C_{ab} x^{a-1} (1-x)^{b-1} dx = 1$$

$$\mu_x = E(x) = \frac{C_{ab}}{C_{a+1,b}} \int_0^1 C_{a+1,b} x^{a+1-1} (1-x)^{b-1} dx$$

x^2 (2)

Poisson process

η unknown, n unknown, $P\eta$ - is known

$\nu \equiv$ intensity
$\equiv$ mean # of counts (events) per unit of t

N_t (# successes by time t)

$N_t \rightleftharpoons \text{Poisson}(\lambda = \nu t)$ amount expected at time t

event $[Y_1 > t] = [N_t = 0]$

$1 - F_Y(t) = P(Y_1 > t) = P[N_t = 0] = P(\text{poisson}(\nu t) = 0)$

$= \frac{(\nu t)^0}{0!} e^{-\nu t}$

$f_Y(t) = \nu e^{-\nu t}$

time increments are indep. and stationary

$\text{mean}_Y = E_Y = \theta = \frac{1}{\nu}$

$f_Y(t) = \nu e^{-\nu t}$ for $t \geq 0$

$= \frac{1}{\theta} e^{-t/\theta}$ for $t \geq 0$

$[W_k > t] = [N_t < k] \Rightarrow P[W_k > t] = P[N_t < k]$

$= \sum_{i=0}^{k-1} P(\text{Poisson}(\nu t) = i)$

$= \sum_{i=0}^{k-1} e^{-\nu t} \frac{(\nu t)^i}{i!}$

derivative of this gives gamma dist.
wait for k hits

Poisson is only continuous distribution that has lack of memory

$P(y > t+s \mid y > s) = \frac{P(y > t+s \text{ and } y > s)}{P(y > s)}$ redundant

$= \frac{e^{-\nu(s+t)}}{\quad} = e^{-\nu t} = P(\ldots)$

4:30 Prob Session
RM C301 Podd

If n gets large, poisson looks normal

Add gammas together starts to look normal

8-4-2009

N_t, $t \geq 0$ is a Poisson process of intensity ν

Then $N_t \cong$ Poisson$(\lambda = \nu t)$ (mean amount in time interval t)

with indep. & stationary increments

Interarrival times $Y_1, Y_2, \ldots$ and Exp [ratio $= \nu$]

pdf $= f_y(y) = \nu e^{-\nu y}$ for $y \geq 0$

Waiting times $W_k = Y_1 + \ldots Y_k$

has Gamma $(k, \text{rate } \nu)$ $\quad f(w) = \frac{\nu^k w^{k-1} e^{\nu w}}{\Gamma(k)}$ for $w \geq 0$

Key fact $[W_t > t] = [N_t < k]$

$[W_k \geq t] \neq [N_t < k]$

$\Rightarrow$

$\not\Leftarrow$

Conditionally uniform

uniformly distributed

ordered statistics (value) put in same order, do not have to arrive in this order

Central limit theorem $\Rightarrow$ applies to distribution with μ and σ^2

Calculating β, Γ moments

$$\int_0^1 C_{a,b} x^{a-1} (1-x)^{b-1} dx = 1$$

$$\mu_k = E(x) = \frac{C_{a,b}}{C_{a+1,b}} \int_0^1 C_{a+1,b} x^{a+1-1} (1-x)^{b-1} dx$$

Independence

Wed C301 3:30

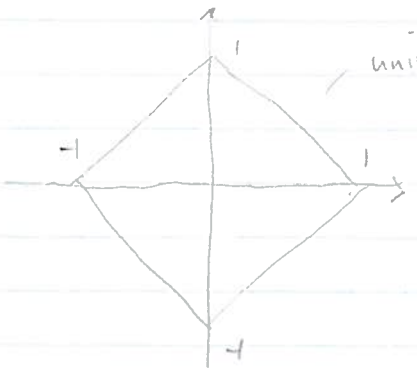
4.9

fudge factor for w/o replacement (for Var) $npq\left[1-\frac{n-1}{N-1}\right]$

Variance of $y = \frac{(1-y)^2}{12}$ uniform

variance of $x = \frac{x^2}{12}$

What would be conditional / marginal distn



uniform over area

14-4-2004 Distributions are in two contexts
Binomial and Bernoulli

Two simultaneous random variables

Marginal distribution

discrete	Cont.
$P(x,y)$	$f(x,y)$

Marginal / conditional fashion is used for models

- Estimated
- Joint probabilities
- Marginal probabilities (located on margins)

Independence $\hat{P}(x,y) \approx \hat{P}(x)\,\hat{P}(y)$

$$\hat{P}_{Y|x}(x) = \hat{P}_{Y|x}(Y|X) = \frac{P_{Y,x}(x,y)}{P_x(x)}$$

Marginal prob cont. hold one variable fixed integrate over other variable
area under curve

- Order statistics
- Poisson Limit Theorem

15-4-2004

What is $\frac{X}{Y=y}$ fix value of y what is $\frac{x}{y}$ distribution

Given y what is x

Density is const along segment → marginal probability

Bernoulli Process (p) Discrete dist interviews

Poisson Process (λ) Continuous dist. watching meteors

Conditional dist

Geometric answers are good

Some math verification

Central Limit Theorem Application approx for Bi with correction

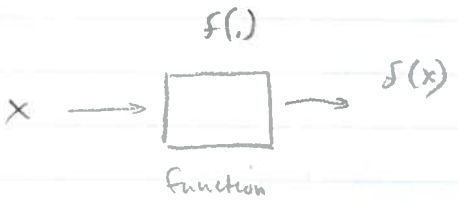
Something combintoral problem

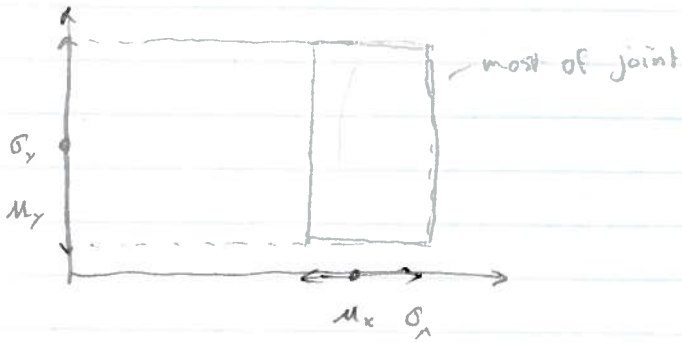
Ch 2, 9, 16, 19 pg 158 Beta

Conditional distributions

where a particular order statistic $X_{n:k}$ (k^{th}) is likely to fall β

Notation





2.1 4-200 Exponential Family

θ - statistical parameter ex: p of Bern

Θ: {0, 1}

Def 1: (Exp Family) Suppose we densities $f_\theta(x)$ [for each $\theta \in \Theta$] of the form:

$$f_\theta(\underline{x}) = c(\theta)\, e^{\left[\sum_{i=1}^{k} w_i(\theta)\, T_i(x)\right]}\, h(x)$$

function of data

E = possibility • probability

Def 2 (canonical form)

$$c(\underset{\sim}{\eta})\, e^{\sum_{i=1}^{k} \eta_i T_i(\underset{\sim}{x})}\, h(\underset{\sim}{x})$$

η_i canonical parameters

Ex: Bernoulli (p) Trials

$$f_p(\underset{\sim}{x}) = \left(\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\right)\left[\prod_{i=1}^{n} 1_{\{0,1\}}(x_i)\right]$$

joint density

indicator function

good for any value of x

$= 0$ if $x \neq 0$ or $x \neq 1$

$$= p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}\left[h(\underset{\sim}{x})\right]$$

$$= \underbrace{(1-p)^n}_{c(p)}\, e^{\underbrace{\sum_{i=1}^{n} x_i}_{T(\underset{\sim}{x})}\underbrace{\left[\log\frac{p}{1-p}\right]}_{w(p)}}\, \underbrace{h(\underset{\sim}{x})}_{h(\underset{\sim}{x})}$$

Odds $\frac{p}{q}$ $\log\left(\frac{p}{q}\right)$ log odds

$$T(x) = \sum_1^n X = \text{Binomial}(n, p)$$

Total num successes

$$f_{T,p}(t) = \binom{n}{t} p^t (1-p)^{n-t}\left[1_{\{0,1,\ldots n\}}(t)\right]$$

$$= p^t(1-p)^{n-t}\left[1_{\{0,1,\ldots,n\}}(t)\right]\underbrace{\binom{n}{t}}_{h(t)}$$

Comment

$$f_{X/T=t}(x) = \frac{p^{\sum x_i}(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}}\left[1_{\left[\sum_i x_i = t\right]} h_1 h_2\right]$$

$$= \frac{1}{\binom{n}{t}}$$

(conditional

t is sufficient for data — Can base anything on it

$$\sigma^2_{x+y} = [(X+Y) - (\mu_x + \mu_y)]^2$$
$$= [(x-\mu_x) + (y-\mu_y)]^2 = \sigma_x^2 + \sigma_y^2 + 2\,\mathrm{Cov}[x,y]$$

B_i ($n=100$, $p=0.75$) same over any

large sigma on denominator mean good approx:

$$f(x,y) = 2e^{-2x} \cdot \frac{1}{x}$$

$f_x(x)$ $f_{Y|X=x}(y)$ ~ uniform $\frac{x}{2}$ could also multiply by x and integrate

$$E\left[\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\ \sigma_y}\right] = \frac{\mathrm{Cov}[x,y]}{\sigma_x \sigma_y} \equiv \rho_{x,y}$$

ρ - can't find nonlinear relationships

26-4-2004

In general

all $a_i = 1$

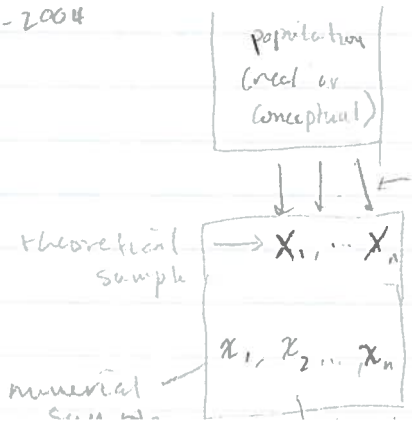
$\max \frac{|a_i - \bar{a}|}{\sqrt{n}\,\sigma_a} \to 0$ as $n \to \infty$ goes to normal

fairly flexible, numbers $1 \to N$ in urn roughly "same size"

$1^2 \to N^2$ " "

quite flexible

28-4-2004



associated with population:

pop. mean μ

pop stdev σ

← n redep. repetitions of experiment X

samples

$\bar{X}_n$ - sample mean or $(\hat{\mu})$

$\hat{s}$ - sample stdev

$s^2 = \frac{1}{n-1}\sum^n (x_i - \bar{x}_n)^2$

Empirical Dist (or this urn test) puts mass $\frac{1}{n}$ at each value $x_1, \ldots, x_n$

this distribution has theoretical $= \bar{X}$

variance $= \hat{\sigma}^2$

$$\text{Var}[X_i] = \text{Var}\left[\frac{T_n}{n}\right] = \frac{1}{n^2} n\sigma_x^2 = \frac{\sigma_x^2}{n}$$ ← gets smaller as $n \to$ large

ES^2 PF] $E\sum_{i=1}^{n}(X - \mu)^2 = E\sum_{i=1}^{n}\left[(X_i - \bar{x}) + (\bar{X} - \mu)\right]^2$

$$= E\sum_{i}^{n}(X_i - \bar{X})^2 + E\sum_{i}^{n}(\bar{X} - \mu)^2$$

$$+ 2(\bar{x} - \mu)\sum(X_i - \bar{X})$$

$$\sum_{i}^{n} x_i - n\frac{\sum x_i}{n}$$

$$= 0$$

Variance of $\bar{X}$

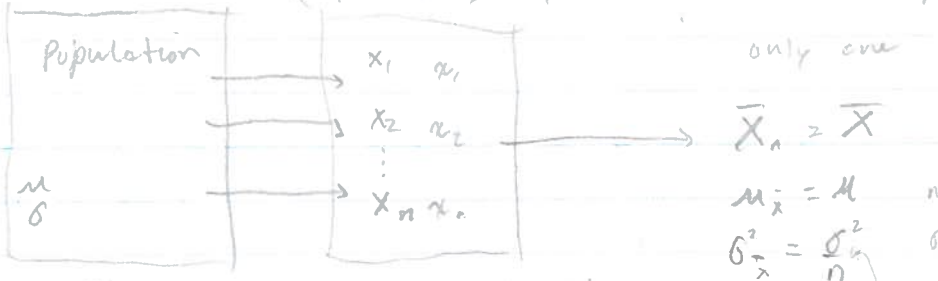
$$E\frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{n\sigma^2 - \sigma^2}{n-1} = \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

$$ES^2 = \sigma^2$$

$$ES \neq \sigma$$

29-4-2009

(model to estimate population parameters by various distributions)



Basic X - experiment Random Sample

$\bar{X}$

$\hat{\sigma}^2$ S^2 ≈ true σ

{emphirical dist}

continuity correction

Continuity approx $P(23 \le X \le 47) \Rightarrow P(22.5 \le X \le 47.5)$

$P(20 \le X)$

$P(19.5 \le X)$ continuity correction

30-4-2004

$$f_{U,V}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{u^2 - 2\rho uv + v^2\}}$$

$Q(u,v) = c^2$, const

given level line

$$h(x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

Indep:

- only of density factors

$\frac{\sigma_y}{\sigma_x}$ - std ev line

Regression only looks at conditional part of a problem

$V = \rho u + \epsilon$ ← natural fluctuation

$Y = m(x) + \epsilon$

$\epsilon \sim N(0, \sigma_\epsilon^2)$

3-5-2004 Normal Theory

Common dist Student's t dist

Def $z \cong N(0,1)$

$u \cong \chi^2_m \equiv$ Chi squared (m)

$v \cong \chi^2_n \equiv$ Chi squared (n)

$$T = \frac{z}{\sqrt{\frac{v}{n}}} \cong T_n$$

Chi squared(m) = Gamma $(\frac{m}{2}, \frac{1}{2})$

$$F = \frac{u/m}{v/n} \cong \mathcal{F}_{m,n}$$

$\frac{u}{u+v} \equiv$ beta (m, n)

SS_{xx} - sum of squares

$$\sum_1^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = N^2(0,1) = \chi^2_n$$

$$= \sum_i^n \left(\frac{X_i - \bar{X} + \bar{X} - \mu}{\sigma}\right)^2 = \sum_i^n \left(\frac{X - \bar{X}}{\sigma}\right)^2 + \frac{n(\bar{X} - \mu)^2}{\sigma}$$

$$\chi^2_{n-1} \qquad \chi^2_1$$

5 may 2024

$Cov[u,v] = Cov[\tilde{u}, \tilde{v}]$ - subtracted mean

$z \approx N(0,1)$
$u \approx \chi^2_m$
$v \approx \chi^2_1$
} ind

$$T = \frac{z}{\sqrt{\frac{v}{n}}} \cong t_{n-1}$$

$$\frac{u/m}{v/n} \cong F_{m,n}$$

$$SS_{xx} = \sum_1^n (X_i - \bar{X})^2$$

u v

assuming $\bar{X}$, X_i are normally distributed

Example: $Cov[\bar{X}, X_i - \bar{X}]$ if $= 0$, shows $\bar{X}$ is ind of all $X_i - \bar{X}, \ldots, X_n - \bar{X}$

Also indep of any function of X_i

$E\ \bar{X}(X_i - \bar{X})$

$$= Cov[\bar{X}, X_i] - Cov[\bar{X}, \bar{X}]$$

$$= Cov\left[\frac{X_1 + X_2 + \cdots X_n}{n}, X_i\right] - Var[\bar{X}]$$

$$= \frac{1}{n}[\sigma_{ii} + \sigma_{2i} + \sigma_{3i} + \cdots \sigma_{ni}] - \frac{\sigma^2}{n}$$

$$= \frac{1}{n}[\sigma^2 + 0 + 0 \cdots 0] - \frac{\sigma^2}{n} = 0 \Rightarrow \text{since normal} \Rightarrow \text{indep}$$

Chi squared $= \frac{SS_{xx}}{\sigma^2} = \sum_1^n \frac{(X_i - \bar{X})^2}{\sigma^2}$

$$\sum_1^n N_i^2(0,1) = \sum_1^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_1^n \left[\left(\frac{X_i - \bar{X}}{\sigma}\right) + \left(\frac{\bar{X} - \mu}{\sigma}\right)\right]^2$$

$$\sum_1^n \chi^2_1 = \sum_1^n \frac{(X_i - \bar{X})^2}{\sigma^2} + \frac{(\bar{X} - \mu)^2}{\sigma^2} + 2 \cdot 0 \quad \text{(indep)}$$

$$\vdots$$

$$\chi^2_n = ? + \chi^2_1$$

$$= ? + \chi^2_1$$

Let $Z_i = \frac{X_i - \mu}{\sigma} \cong N(0,1)$, IND

$Z = \sqrt{n}\,\bar{Z}$ $\qquad \frac{SS_{xx}}{\sigma^2} = \sum_1^n (Z_i - \bar{Z})^2$ $\qquad \underset{\sim}{Z} \equiv \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$

$$\sum_i^n Z_i^2 = \underset{\sim}{Z}'\underset{\sim}{Z} = [Z_1 \dots Z_n]\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

$$= \frac{SS_{xx}}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = \underbrace{\sum_2^n (Z_i^2 \text{ residual about fit})}_{\chi^2_{n-1}} + Z^2$$

n-1 directions left over, live in n-1 dimension, sum to zero (one side condition)
Residuals will look like χ^2, after fitting perpendicular to fit
will have n-1 wiggle space (fluctuations) n-1 d.o.f.

$$\frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma} = N(0,1) \qquad T = \frac{\sqrt{n}\,(\bar{x} - \mu)}{s} \cong T_{n-1}$$

$$\frac{SS_{xx}}{\sigma^2} \cong \chi^2_{n-1}$$

normal ← $\chi^2_n = \sum_1^n Z_i = Z_1^2 + \sum_2^n Z_i^2$ → chi squared n-1

6-MAY-2004 Ch 7-6

7-MAY-2004

Model eqn $Y_i = \mu + \varepsilon_i$

LSE $SS(a) = \sum_i^n (y_i - a)^2$ minimize sum of squares

How good is this?

α - over center point $\qquad \gamma$ - over baseline

Ch 7 Estimation I

Def A statistic $T = T(\underset{\sim}{x})$ some function of data
a computable function of the data

$t = \frac{\sqrt{n}\,(\bar{x} - \mu)}{s}$ μ not computable
not a statistic

Sub $x_i = \mu + \varepsilon_i$ and find confidence interval $\mu = \bar{X} \pm t^* \frac{s}{\sqrt{n}}$

The set Θ of all possible values θ of our parameter is called the parameter space. Has parameter space P_θ $(0,1)$ $(0,\infty)$, $(-\infty,\infty)\times(0,\infty)$

What do we care about:

may care about θ, $\mu + 1.96$

" $g(\theta)$ ($\sqrt{pq}$), some function of θ

(1) If $E_\theta T(\underline{X}) = g(\theta)$ for all of Θ, call T an unbiased estimator of $g(\theta)$

$\hat{p} = \bar{X}$ $E_p \bar{X} = p$ $\bar{X}$ on average gives p

accuracy precision

The quantity $B(\theta) = \text{Bias}(\theta) = E_\theta\{T(\underline{X})\} - g(\theta)$

" " MSE - mean squared error $(\theta) = E_\theta\left(T(\underline{X}) - g(\theta)\right)^2$

How much did it miss true value

on average what is squared error

If unbiased $= \text{Var}_\theta[T(\underline{X})]$

$$= E_\theta\left\{\left[T(\underline{X}) - E_\theta T\right] + \left[E_\theta(T) - g(\theta)\right]\right\}^2$$

$$= E_\theta\left\{\left(T(\underline{X}) - E_\theta T\right)^2\right\} + \text{Bias}^2(\theta) + 2\left(E_\theta T - g(\theta)\right)\left[E_\theta T - E_\theta T\right]$$

$\left[E_\theta T - E_\theta T\right] = 0$

11-5-2004

$P = \{P_\theta : \theta \in \Theta\}$ { a family of distributions }

Model: $\underline{X} \sim P_\theta$ for some unknown $\theta \in \Theta$

" we want use to use $X_1, \ldots, X_n$ to est which θ is true

Generally, we want to use $T(\underline{X})$ to est $g(\theta)$

$\text{Bias}(\theta) = E_\theta T(\underline{X})$ ← what is average of T when θ is θ

$\text{Bias}(\theta) = E_\theta T(\underline{X}) - g(\theta)$, MSE

Two methods to estimate error

MME, MLE

- Method of moments.
  - P_θ has mean μ_θ and 2nd moment $E_\theta X^2$
  - Data $\underline{x}$ (or $\underline{X}$) has mean $\bar{x}$ (or $\bar{X}$) and 2nd moment $+ \bar{X}^2$
    $= \frac{1}{n}\sum_i x_i^2$

Today, 1

13-5-2024

$$l_\theta = n\bar{x}\log\theta + n(1-\bar{x})\log(1-\theta) + \log h(x)$$

$$0 = \dot{l}_\theta = \frac{n\bar{x}}{\theta} - \frac{n(1-\bar{x})}{1-\theta} = \frac{n\bar{x} - n\bar{x}\theta - \theta n + n\bar{x}\theta}{\theta(1-\theta)}$$

$0 = \frac{(\bar{x}-\theta)n}{\theta(1-\theta)}$ solution is: $\hat{\theta}$: $\bar{x} - \theta = 0 \Rightarrow \bar{x} = \theta = \hat{\theta}_n$

$\frac{n}{\theta(1-\theta)}(\bar{x}-\theta)$

↳ sigma

Binomial $L_x(\theta) = \binom{n}{x}\theta^x(1-\theta)^{1-x}\, 1_{\{0,\dots,n\}}(x)$

Lets keep in mind the MME $\bar{\theta}_n = \theta$

Since $\mu_x = \frac{\theta}{2}$ (by MME principle) $\bar{X}$, Thus $\bar{\theta}_n = 2\bar{X}$

- Note that $\sqrt{n}(\bar{\theta} - \theta) = 2\sqrt{n}(\bar{x} - \mu_x) \xrightarrow{\text{converge}} N(0, 4\sigma_x^2) = N(0, 4\cdot\frac{\theta^2}{12}) = N(0, \theta^2/3)$
- No to analyze T

Start $F_T(t) = P(T \le t) = P(\max X_i \le t) = P^n(x_1 \le t) = F_X^n(t) = \left(\frac{t}{\theta}\right)^n$ on $0 \le x \le \theta$

$E_\theta \hat{\theta}_n = E_\theta T = \frac{n}{n+1}\theta$ Let's use $\tilde{\theta}_n = \frac{n+1}{n}(T)$ is unbiased

$Var_\theta[\tilde{\theta}_n] = \left(\frac{n+1}{n}\right) Var_\theta[T] = \frac{\theta^2}{n(n+2)}$ super efficient

$P_\theta[n(\theta - \hat{\theta}_n) > t)$

$= P_\theta(\hat{\theta}_n < \theta - \frac{t}{n}) = P_\theta(T < \theta - \frac{t}{n}) = P_\theta\left(\frac{T}{\theta} < 1 - \frac{t/\theta}{n}\right) = P(u_{min} < 1 - \frac{t/\theta}{n})$

$= (1 - \frac{t/\theta}{n})^n \to e^{-t/\theta} = P(Exp(mean = \theta) > t)$

14-5-2024

$N = Y_1 + Y_2 \cdots Y_{10}$ ↖ whether not

$N = X_1 + X_2 + \cdots X_5$

Beneath:

$\mu_N = 10\,\mu_Y = 10\,P_Y$ (10 pairs; 10 times itself; 90 times others)

$\sigma_N = 10\,\sigma_Y^2 + (10\cdot 9)\,Cov[Y_1, Y_2]$ (cross products terms)

unknowns P_Y, $E(Y_1Y_2)$ (1st and 2nd R B; 1st B 2nd R)

$= 10\,P_Y(1-P_Y) + 10[E(Y_1Y_2) - P_Y P_Y]$

$P_Y = \frac{C_1^5 C_1^5}{C_2^{10}} = \frac{5\cdot 5}{55}$

all means the same
all variances same

$$E(Y_1 Y_2) = P(Y_1 = 1 \text{ and } Y_2 = 1) = P(Y_1 = 1) P(Y_2 = 1 \mid Y_1 = 1) = P_y \cdot \frac{C_1^4 C_1^4}{C_2^{18}}$$

why? chose second

$P_x = \frac{5}{14}$ $\quad E(X_1 X_2) = \frac{20}{20}, \frac{18}{20}, \frac{5}{18}, \frac{4}{17}$

where can R_1 live?
where does R_2 live?

R_1's friend $\quad R_2$'s friend

EX 8

X_i: ξ_0'	Y_i: ξ_0'	Z_i: ξ_0'
θ^2	$2\theta(1-\theta)^2$	$(1-\theta)^2$
P_A	P_B	P_C

3 groups

One observation:

$$P_A^{X_i} P_B^{Y_i} P_C^{Z_i}$$

For many:

$$\prod_i P_A^{X_i} P_B^{Y_i} P_C^{Z_i} = P_A^{\sum X_i} P_B^{\sum Y_i} P_C^{\sum Z_i}$$

$$= P_A^{N_A} P_B^{N_B} P_C^{N_C}$$

$$= \theta^{2N_A} (2\theta(1-\theta)^2)^{N_B} \cdot (1-\theta)^{2N_C}$$

Do §§

16-MAY-2004

Ex 7 $Y_1, \ldots, Y_n$ be ind $\quad Y_i = N(\mu_i(\theta), \sigma^2) \qquad \theta = (\alpha, \beta)$

regression $\mu_i(\theta) = \alpha + \beta(x_i - \bar{x})$

One Sample Problem

Sum of Squares $= H_{\underset{\sim}{y}}(\theta) = \sum_i (y_i - \mu)^2$

min by $\hat{\mu} = \hat{\theta} = \bar{y}$

Two Sample Problem:

$$H_{\underset{\sim}{y}}(\theta) = \sum_i (x_i - \mu)^2 + \sum_{i=1} (y_i - \nu)^2$$

mean by $\hat{\mu} = \bar{x}$ and $\hat{\nu} = \bar{y}$

Regression model

$$H_{\underset{\sim}{y}}(\theta) = \sum_i (y_i - (\alpha + \beta(x_i - \bar{x}))^2$$

need to min

$\hat{\alpha} = \bar{y}$, $\hat{\beta} = \frac{\sum (x_i - \bar{x})}{S_{xx}}$

19-MAY-2004 7.23, 7.24

$\int_0^\infty x^2 e^{-2x}\,dx \Rightarrow$ Examine $-\frac{1}{2}x^2 e^{-2x}$

derivative $--\frac{1}{2}x^2 2e^{-2x}$

if $\int_0^\infty x^2 e^{-2x}\,dx$

↓ make look like $F(\alpha)$

$$f_\theta(y) = \frac{y}{\theta^2} e^{-\frac{y^2}{2\theta^2}}$$

(5) $L(y)_\theta$ $\quad L_\theta(y) = \prod_{i=1}^{n} \frac{y_i}{\theta^2} e^{-\frac{y_i^2}{2\theta^2}} = + \frac{\prod y_i}{\theta^{2n}} e^{-\frac{\sum y_i^2}{2\theta^2}}$

$$l_\theta = \log \frac{y_i}{\theta^{2n}} + \frac{-\sum y_i^2}{2\theta^2} = -2n \log\theta - \frac{n\overline{y^2}}{2\theta^2}$$

$$0 = \dot{l}_\theta = \frac{\partial}{\partial\theta} l_\theta = -\frac{2n}{\theta} + \frac{2n\overline{y^2}}{2\theta^3} = \frac{n}{\theta^3}\left[\overline{y^2} - 2\theta^2\right] \Rightarrow \hat{\theta} = \sqrt{\tfrac{1}{2}\overline{y^2}}$$

↑ can't be zero

(2)

$$\hat{m} = \frac{\bar{x}_{2n} + 3\bar{y}_{3n}}{4} \qquad \frac{m - 3m + 3(m+\gamma)}{4} = \frac{4m}{4} = m$$

$$\mathrm{Var}[\hat{m}] = \frac{1}{16}\left\{\frac{2\sigma^2}{2n} + 3^2\frac{3\sigma^2}{3n}\right\} = \frac{\sigma^2}{n}\left(\frac{10}{16}\right) = \frac{5}{8}\frac{\sigma^2}{n}$$

$$\xi = \frac{\sqrt{n}(\hat{m} - m)}{\sqrt{\frac{5}{8}}\,\sigma} \cong N(0,1)$$

$$SS_{xx} \cong 2\sigma^2 \chi^2_{2n-1} = \sum_1^{2n}(x_i - \bar{x})^2$$

$$SS_{yy} \cong 3\sigma^2 \chi^2_{3n-1} = \sum_1^{3n}(y_i - \bar{x})^2$$

can't add directly, take $\frac{1}{2}$ of top and $\frac{1}{3}$ of bottom one

$$c\sigma^2\left(\chi^2_{2n-1} + \chi^2_{3n-1}\right) = c\sigma^2\chi^2_{5n-2} \qquad c \text{ const}$$

Use $S_p^2 = \frac{\frac{1}{2}S_{xx} + \frac{1}{3}S_{yy}}{5n-2} = \frac{\sigma^2 \chi^2_{5n-2}}{5n-2}$

$T = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\frac{5}{8}}\, S_p} = t_{5n-2}$

$\Rightarrow$ CI for μ is $\hat{\mu} \pm t^* \frac{1}{\sqrt{n}} \sqrt{\frac{5}{8}}\, S_p$

$S_p^2 = \frac{\frac{1}{2}SS_{xx} + \frac{1}{3}SS_{yy}}{5n-2} = \frac{2n-1}{5n-2}\frac{\frac{1}{2}S_{xx}}{2n-1} + \frac{3n-1}{5n-2}\frac{\frac{1}{3}S_{yy}}{3n-1}$

$= c\,S_x^2 + (1-c)S_y^2$

- weighted linear combo of two variances
- $SS_{xx} = \sum_1^n (X_i - \bar{X})^2$
- $S_x^2 = \frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2 = \frac{SS_{xx}}{n-1}$

(#3) $M_y(\theta) = \bar{y}_{data}$ then solve for θ

theoretical mean

$= E_\theta y = \int_0^\infty y \frac{y}{\theta^2} e^{-\frac{y^2}{2\theta^2}} dy$ use $z = \frac{y^2}{2\theta^2}$

$= \int_0^\infty \frac{2y}{2} \frac{y}{\theta^2} e^{-\frac{y^2}{2\theta^2}} dy$ try to make look like $\Gamma(\alpha)$

$dy = \frac{\theta}{\sqrt{2z}} dz$

$= \int_0^\infty \frac{2\theta}{\sqrt{2z}} z\, e^{-z} dz$

$= \sqrt{2}\theta\Gamma(\tfrac{3}{2}) \int_0^\infty \underbrace{\frac{1}{\Gamma(\frac{3}{2})} z^{\frac{3}{2}-1} e^{-z}}_{=1} dz = \sqrt{2}\theta \tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \frac{\sqrt{2}\theta\sqrt{\pi}}{\sqrt{4}}$

MME $= \sqrt{\frac{\pi}{2}}\,\hat{\theta} = \bar{y} \Rightarrow \hat{\theta} = \sqrt{\frac{2}{\pi}}\,\bar{y}$

Find variance σ_y^2 $\quad$ LWSE $\sqrt{n}(\hat{\theta} - \theta) \to N(0, \frac{2}{\pi}\sigma_y^2)$ by Cent. Limit Thm.

(#1) $F_\theta(y) = \int_0^y f_\theta(y')dy' = \int_0^{\frac{y^2}{2\theta^2}} e^{-z} dz = 1 - e^{-\frac{y^2}{2\theta^2}}$

$\left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\theta}}\right]^{2n} e^{-\frac{1}{2(2\theta^2)} \sum_1^{2n} (x_i - (\mu - 3\nu))^2} \times$

$$\times \left[\frac{1}{\sqrt{2\pi}\sqrt{36^2}}\right]^{3n} e^{-\frac{1}{2(36^2)}\sum_1^{3n}(y_i-(\mu+\nu))^2}$$

$\ell = \log$

20-5-2004

$$\text{Var}[\hat{\theta}] = \frac{\sigma_1^2}{n} \qquad \text{Var}[\tilde{\theta}] = \frac{\sigma_2^2}{m}$$

$\frac{n}{m} = \frac{\sigma_1^2}{\sigma_2^2}$ ratio of samples sizes to do equally well ⇒ called efficiency

F dist are estimated variances

Bayes estimators

model : The basic RV exp has dist $f_\theta(x) = f(x|\theta)$ where the value θ of the random var Θ is dist. according to the prior dist $\pi(\cdot)$ So we obtain infor by observing data $\underset{\sim}{X}$ according to $f_\theta(\underset{\sim}{x})$. Then $\pi(\theta|x) = \frac{f(x,\theta)\pi(\theta)}{\int_{\text{all }\theta'} f(x,\theta')\pi(\theta')\,d\theta'}$ (θ' dummy variable)

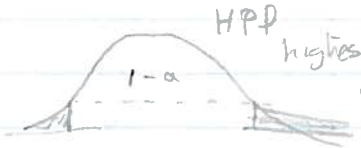
$$= \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{f_{\underset{\sim}{X}}(\underset{\sim}{x})} = \frac{P(B|A_i)P(B_i)}{\sum_{\text{all } i} P(B|A_i)P(A_i)}$$

• All actions will be taken on basis of posterior dist $\pi(\cdot|\underset{\sim}{x})$

(3) $m(x) = E(\theta|\underset{\sim}{x})$ would seem to be a reasonable estimator of θ mean of posterior dist

$\text{Var}(\theta|\underset{\sim}{x})$ would seem to be useful in assessing the value of the Bayes est $E(\theta|\underset{\sim}{x})$

Bayesian interval contains most of probability of



very likely points give short interval

$$\pi(\theta) \xrightarrow{\text{update}} \pi(\theta|\underset{\sim}{x}) \to \pi(\theta|\underset{\sim}{x},\underset{\sim}{y})$$

well designed experiment σ is small

Ex 1: $X_1, \ldots, X_n$ Bernoulli(θ)

$\pi(\theta) \sim$ Beta(α, β)

$$\pi(\theta) = \left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \right) \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\pi(\theta|x) = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}\,C_{\alpha,\beta}}{f_x(\underset{\sim}{x})} \qquad \text{let } y = \sum_1^n x_i$$

($C_{\alpha,\beta}$: const)

$$= \theta^{(y+\alpha-1)}(1-\theta)^{(n+\beta-y-1)}\,C_{\alpha,\beta,\underset{\sim}{x}} \qquad \text{conjuge priod}$$

$$= \text{Beta}(\alpha+y, n+\beta-y)$$

21-5-2004

$l_\theta(x) = \log f_\theta(x)$

$$\log \prod_1^n f_\theta(X_i) = l_\theta(\underset{\sim}{x})$$
$$\sum \log f_\theta(X_i)$$
$$= \sum_i l'_\theta(x_i)$$

$i_\theta = \int_\theta \log f_\theta(X_i)$ has mean zero

$$\sqrt{n}(\hat{\theta}-\theta) \Rightarrow N(0, 1^{-1}(\theta))$$

26-5-2004

Joint . conditional

completeness T - function of data

$g(T)$ - any function of T

$f_\theta(\,)$ some dist. , smart or stupid guess give same result because of completeness

Find an unbiased estimator, then one is done

- Fails when have nontrivial eqn

27-5-2004 $\left\{ \left(\frac{\mu+\nu}{\sigma^2}, \frac{2\mu+\nu}{\sigma^2}, -\frac{1}{2\sigma^2} \right) : \begin{matrix} \sigma^2 > 0 \\ -\infty \le \mu, \nu \le \infty \end{matrix} \right\}$

3 coord vector is sufficient for 3 curved statistic

$\sum_{i=1}^m X_i = \bar{X}_m$ knowing sum is same as knowing ave

or $T(\bar{X}_m, \bar{Y}_n, S_p)$ is also sufficient & complete

26-5-2004 - Power(μ) = 1 - P(Type II)

©1996 by Yves Nievergelt. To appear in *Encyclopedia of Statistical Sciences.*

TOTAL LEAST SQUARES

Total Least Squares (TLS) fits a hyperplane to data with errors in *all* their coordinates.

Historically, Van Huffel & Vandewalle [17] credit R. J. Adcock [1] for the first publication, in 1878, about fitting lines to data in the plane through TLS. They also credit G. H. Golub & C. F. Van Loan [5] for solving the TLS problem through the Singular-Value Decomposition " (SVD) [14] and thus providing extensively documented theory and algorithms for TLS. Numerical methods had been used earlier [2], but without a complete theoretical justification.

Total Least Squares admits several mathematically equivalent formulations:

(GTLS) Geometric: to m data points $\vec{c}_1, \ldots, \vec{c}_m$ in $\mathbb{R}^{n+1}$, fit a hyperplane $H \subset \mathbb{R}^{n+1}$ with equation $\langle \vec{c} - \vec{c}_0, \vec{x} \rangle = 0$, where $\langle \ , \ \rangle$ denotes the dot product, by finding a point $\vec{c}_0 \in H$ and a non-zero perpendicular vector $\vec{x} \perp H$ that minimize the sum D of the squared distances measured perpendicularly to the hyperplane from each data point:

$$D(\vec{c}_1, \ldots, \vec{c}_m; \vec{c}_0, \vec{x}) = \sum_{i=1}^{m} \frac{|\langle \vec{c}_i - \vec{c}_0, \vec{x} \rangle|^2}{\langle \vec{x}, \vec{x} \rangle}.$$

(ATLS) Algebraic: for a linear system $A\vec{z} = \vec{b}$, which need not have a solution if the columns of $C = [A; \vec{b}]$ are linearly dependent, determine a solution $\vec{z}$ of a related linear system $\hat{A}\vec{z} = \hat{\vec{b}}$, with $\hat{A}$ and $\hat{\vec{b}}$ minimizing the sum of the squares of all the entries in $\vec{b} - \hat{\vec{b}}$ and $A - \hat{A}$. Equivalently, TLS determines $\hat{C} = [\hat{A}; \hat{\vec{b}}]$ with linearly dependent columns minimizing the Frobenius norm $\|S\|_F^2 = \sum_{i,j} s_{i,j}^2$ of $S = C - \hat{C}$.

(STLS) Statistical: to m data points $\vec{c}_1, \ldots, \vec{c}_m$ in $\mathbb{R}^{n+1}$, fit a general linear model " $\langle \vec{c} - \vec{c}_0, \vec{x} \rangle = 0$ under the hypothesis that *all* the coordinates of all data come from statistically independent " standard " normal " random variables (Errors-In-Variables model).

Despite the *linearity of the model* equation $\langle \vec{c} - \vec{c}_0, \vec{x} \rangle = 0$, the TLS problem leads to *non-linear equations* for the minimizing model parameters $\vec{x}$. Yet all three formulations are mathematically equivalent, with the data points $\vec{c}_1, \ldots, \vec{c}_m$ in $\mathbb{R}^{n+1}$ playing the role of the rows of $C = [A; \vec{b}]$, and with Euclidean distance corresponding to standard deviation ". Consequently, all three problems lend themselves to a solution through a common algorithm, developed by Golub & Van Loan [5]. [7] and extended by Van Huffel & Vandewalle [17], and which consists essentially in a SVD of C. The SVD of the matrix $C \in \mathbb{M}_{m \times (n+1)}(\mathbb{R})$ with m rows and $(n+1)$ columns of entries in the reals $\mathbb{R}$ has the form $C = U \cdot \Sigma \cdot V'$ consisting of a diagonal matrix $\Sigma \in \mathbb{M}_{m \times (n+1)}(\mathbb{R})$ and orthogonal " matrices $U \in \mathbb{M}_{m \times m}(\mathbb{R})$ and $V \in \mathbb{M}_{(n+1) \times (n+1)}(\mathbb{R})$, with the superscript $'$ denoting transposition. The theory, algorithms, and analysis of computers' round-off " errors for the SVD have been extensively researched and documented [3], [4], [7], [10], [15], [18], [19]. Thus, it is the SVD that yields the theoretical and globally convergent algorithmic solution of the TLS.

The concept of total least squares regression also applies to the fitting of surfaces other than hyperplanes, defined by non-linear equations $f(\vec{c}; \vec{x}) = 0$. The resulting problem of Nonlinear TLS consists in determining the model's parameters $\vec{x}$ that minimize $\sum_i [f(\vec{p}_i; \vec{x})]^2$. In special situations, for instance, in fitting a circle or sphere to data points [13], or in fitting parameters to special chemical equations [11], Nonlinear TLS may admit a solution in closed form, but in general no globally convergent algorithm exists to compute $\vec{x}$, in contrast to linear TLS with SVD.

Solution of TLS through SVD

Given m data $\vec{c}_i \in \mathbb{R}^{n+1}$ the problem of TLS consists in finding $\vec{c}_0, \vec{x} \in \mathbb{R}^{n+1}$ minimizing

$$D(\vec{c}_1, \ldots, \vec{c}_m; \vec{c}_0, \vec{x}) = \sum_{i=1}^{m} \frac{|\langle \vec{c}_i - \vec{c}_0, \vec{x} \rangle|^2}{\langle \vec{x}, \vec{x} \rangle}.$$

A simplification arises from locating $\vec{c}_0$ at the sample mean " $\bar{\vec{c}} = (1/m) \sum_{i=1}^{m} \vec{c}_i$, because $D(\vec{c}_1, \ldots, \vec{c}_m; \vec{c}_0, \vec{x}) \geq D(\vec{c}_1, \ldots, \vec{c}_m; \bar{\vec{c}}, \vec{x})$ for all data $\vec{c}_i$ and all parameters $\vec{c}_0$ and $\vec{x}$. To find $\vec{x}$, observe that $D(\vec{c}_1, \ldots, \vec{c}_m; \bar{\vec{c}}, \vec{x}) =$

©1996 by Yves Nievergelt. To appear in *Encyclopedia of Statistical Sciences*.

$\|\vec{z}\|_2^2/\|\vec{x}\|_2^2 = \|C\cdot\vec{x}\|_2^2/\|\vec{x}\|_2^2$, where $\|\vec{u}\|_2 = \sqrt{\langle\vec{u},\vec{u}\rangle}$ denotes the Euclidean norm and $C \in \mathbb{M}_{m\times(n+1)}(\mathbb{R})$ denotes the matrix with i-th row $\vec{c}_i - \overline{\vec{c}}$:

$$C = \begin{pmatrix} \vec{c}_1 - \overline{\vec{c}} \\ \vdots \\ \vec{c}_m - \overline{\vec{c}} \end{pmatrix}.$$

Consequently, D reaches its minimum at a unit vector $\vec{v} = \vec{x}/\|\vec{x}\|_2$ that minimizes $\|C\cdot\vec{v}\|_2$. The theory of the SVD shows that $\vec{v}$ coincides with any singular vector $\vec{v}_{n+1}$ (eigenvector* of $C'C$) corresponding to the smallest singular value σ_{n+1} of C, where $\sigma_1^2 \geq \cdots \geq \sigma_{n+1}^2 \geq 0$ represent the eigenvalues* of $C'C$. With $\vec{x} = \vec{v}_{n+1}$ and σ_{n+1} computed, $\widehat{\vec{c}}_i = \vec{c}_i - \langle\vec{c}_i - \overline{\vec{c}}, \vec{v}_{n+1}\rangle\vec{v}_{n+1}$ represents the orthogonal projection of the data $\vec{c}_i$ onto H:

$$\widehat{C} = \begin{pmatrix} \widehat{\vec{c}}_1 \\ \vdots \\ \widehat{\vec{c}}_m \end{pmatrix} = \begin{pmatrix} \vec{c}_1 \\ \vdots \\ \vec{c}_m \end{pmatrix} - (M\cdot\vec{v}_{n+1})\cdot\vec{v}'_{n+1}$$

and $D(\vec{c}_1,\ldots,\vec{c}_m;\overline{\vec{c}},\vec{v}_{n+1}) = \sum_{i=1}^m \|\widehat{\vec{c}}_i - \vec{c}_i\|_2^2 = \|C\cdot\vec{v}_{n+1}\|_2^2 = \sigma_{n+1}^2$.

Statistical Formulation of TLS

The statistical formulation of TLS consists in fitting a linear model $\langle\vec{c}-\vec{c}_0,\vec{x}\rangle = 0$ to data $\vec{c}_1,\ldots,\vec{c}_m \in \mathbb{R}^{n+1}$, where *all* coordinates of each $\vec{c}_i - \vec{c}_0$ represent realizations of independent standard normal random variables (after a rescaling of the initial data by the measured covariance matrix if necessary [16]). Specifically, the statistical model assumes that there exist a fixed vector $\vec{c}_0 \in \mathbb{R}^{n+1}$ and a non-zero vector of the model's parameters $\vec{x} = (x_1,\ldots,x_{n+1}) \in \mathbb{R}^{n+1}$, such that each vector $\vec{c}_i$ of observations decomposes as a sum

$$\begin{aligned} \vec{c}_i &= \vec{c}_0 + \vec{w}_i + \vec{r}_i, \\ \vec{c}_0 &\in \mathbb{R}^{n+1}, \\ \langle\vec{w}_i,\vec{x}\rangle &= 0, \end{aligned}$$

with a vector $\vec{w}_i$ satisfying the linear equation $\langle\vec{w}_i,\vec{x}\rangle = 0$, and a vector of "residuals" $\vec{r}_i$ where all coordinates come from statistically independent normal random variables with mean equal to zero and with common variance σ^2. The TLS solution in the preceding section then corresponds to a minimum variance estimator* of the model's parameters $\vec{x}$, which seeks $\vec{x}$ and estimated vectors $\widehat{\vec{c}}_0,\widehat{\vec{c}}_1,\ldots,\widehat{\vec{c}}_m$ such that $\langle\widehat{\vec{c}}_i - \widehat{\vec{c}}_0,\vec{x}\rangle = 0$ for every i so as to minimize $\sum_{i=1}^m \|\vec{c}_i - \widehat{\vec{c}}_i\|_2^2$. The resulting estimators from TLS then coincide with the consistent* and unbiased* sample mean [9], $\widehat{\vec{c}}_0 = \overline{\vec{c}} = (1/m)\sum_{i=1}^m \vec{c}_i$, and with the TLS solution $\widehat{\vec{x}} = \vec{v}_{n+1}$ of $C\cdot\vec{x} = \vec{0}$, where $\vec{x}$ has unit Euclidean length and the i-th row of C contains $\vec{c}_i - \overline{\vec{c}}$, as in the geometric and algebraic formulations. Moreover, the square of the smallest singular value σ_{n+1}^2 of C provides the resulting estimator $\widehat{\sigma^2} = (1/m)\sigma_{n+1}^2$ for the common variance σ^2. Finally, $\widehat{\vec{c}}_i = \vec{c}_i - (C\cdot\vec{v}_{n+1}\cdot\vec{v}'_{n+1})_i$ estimates the corresponding point $\vec{w}_i + \overline{\vec{c}}$ on the linear model, where $(C\cdot\vec{v}_{n+1}\cdot\vec{v}'_{n+1})_i$ denotes the i-th row of the $m\times(n+1)$ matrix $C\cdot\vec{v}_{n+1}\cdot\vec{v}'_{n+1}$.

In contrast to Ordinary Least Squares* (OLS), the TLS estimators $\widehat{\vec{x}}$, $(1/m)\sigma_{n+1}^2$, $\overline{\vec{c}}$, and $\widehat{\vec{c}}_i$ are *consistent* maximum likelihood* estimators of all of $\vec{x}$, σ^2, $\vec{c}_0$, and $\vec{c}_i$ [17]. Therefore, TLS appears to yield better estimates of the model's parameters $\vec{x}$ [17]. However, with a variable x_{n+1} distinguished as depending upon the others, $x_1,\ldots,x_n$, then OLS appears to provide better predictions of c_{n+1} in terms of $c_1,\ldots,c_n$ [17].

Example

Consider the five data points in space displayed in figure 1,

$$\begin{aligned} \vec{c}_1 &= (\ 43,\ 15,\ 127\ ) \\ \vec{c}_2 &= (\ 67,\ 135,\ 13\ ) \\ \vec{c}_3 &= (\ 11,\ 125,\ 69\ ) \\ \vec{c}_4 &= (\ 54,\ 40,\ 146\ ) \\ \vec{c}_5 &= (\ 75,\ 85,\ 95\ ) \end{aligned}$$

©1996 by Yves Nievergelt. To appear in *Encyclopedia of Statistical Sciences.*

Calculate the sample mean vector $\bar{\vec{c}} = (1/5)\sum_{i=1}^{5} \vec{c}_i = (50, 80, 90)$, and form the matrix

$$C = \begin{pmatrix} \vec{c}_1 - \bar{\vec{c}} \\ \vec{c}_2 - \bar{\vec{c}} \\ \vec{c}_3 - \bar{\vec{c}} \\ \vec{c}_4 - \bar{\vec{c}} \\ \vec{c}_5 - \bar{\vec{c}} \end{pmatrix} = \begin{pmatrix} -7 & -65 & 37 \\ 17 & 55 & -77 \\ -39 & 45 & -21 \\ 4 & -40 & 56 \\ 25 & 5 & 5 \end{pmatrix}.$$

Compute the smallest singular value of C, here $\sigma_{n+1} = \sigma_3 = 30$, and the corresponding singular vector, $\vec{v}_{n+1} = (1/3, 2/3, 2/3)'$. Thus, the hyperplane H passes through $\vec{c}_0 = \bar{\vec{c}} = (50, 80, 90)$ and lies perpendicularly to the vector $\vec{x} = \vec{v}_{n+1} = (1/3, 2/3, 2/3)'$, so that H satisfies the equation

$$(1/3)(x - 50) + (2/3)(y - 80) + (2/3)(z - 90) = 0.$$

Moreover, $C \cdot \vec{v}_{n+1} = (-21, -9, 3, 12, 15)'$ contains the signed distances $d(\vec{c}_i, H)$ from the data points to the hyperplane H, measured parallelly to $\vec{v}_{n+1}$, so that $\sum_{i=1}^{5} d(\vec{c}_i, H)^2 = \|C \cdot \vec{v}_{n+1}\|_2^2 = \sigma_{n+1}^2 = 30^2$, which gives the orthogonal projections $\hat{\vec{c}}_1, \ldots, \hat{\vec{c}}_5$ of the data on H:

$$\hat{C} = \begin{pmatrix} \hat{\vec{c}}_1 \\ \vdots \\ \hat{\vec{c}}_m \end{pmatrix} = \begin{pmatrix} \vec{c}_1 \\ \vdots \\ \vec{c}_m \end{pmatrix} - \begin{pmatrix} -21 \\ -9 \\ 3 \\ 12 \\ 15 \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 & 2/3 \end{pmatrix} = \begin{pmatrix} 50 & 29 & 141 \\ 70 & 141 & 19 \\ 10 & 123 & 67 \\ 50 & 32 & 138 \\ 70 & 75 & 85 \end{pmatrix}.$$

Statistically, the sample mean $\bar{\vec{c}} = (50, 80, 90)$ estimates the constant $\vec{c}_0$, and the square of the smallest singular value of C estimates the variance, $\widehat{\sigma^2} = (1/m)\sigma_{n+1}^2 = (1/5)30^2 = 180$. Similarly, the singular vector $\vec{v}_{n+1}$ corresponding to σ_{n+1} estimates the model's parameters, $\hat{\vec{x}} = \vec{v}_{n+1} = (1/3, 2/3, 2/3)'$. Finally, the adjusted vectors $\hat{\vec{c}}_i$ in example 1 estimate the "true" values of the observations $\vec{c}_i$, for $i \in \{1, \ldots, 5\}$, for instance, $\hat{\vec{c}}_1 = (50, 29, 141) \in H$ See also [12].

Extensions

The generic TLS just presented admits several extensions: to multiple right-hand sides with $AX = B$ and $C = [A; B]$, to multiple solutions with repeated smallest singular values $\sigma_{p+1} = \cdots = \sigma_{n+1}$, to alternate column selections if $(\vec{v}_{n+1})_{n+1} = 0$, and to "mixed" problems with some columns of C known exactly and other columns adjusted through TLS [17].

Computations

Real TLS problems require extensively developed and tested scientific software, for instance, the `PTLS` routine in the `VANHUFFEL` directory at `netlib@research.att.com` on the Internet [8].

Acknowledgement

This work was supported in part by the National Science Foundation's grant DUE-9255539.

References

[1] Adcock, R. J. (1878). "A Problem In Least Squares." *The Analyst*, **5**, 53–54. (Earliest contribution to TLS cited in [17].)

[2] Deming, W. E. (1943). *Statistical Adjustment of Data*, Wiley, Chichester, UK. Reprinted (1968), Dover, New York, NY. (Description of earlier numerical TLS on p. 153.)

[3] Demmel, J., and Kahan, W. (1990). "Accurate singular values of bidiagonal matrices." *SIAM J. Sci. Stat. Comp.*, **11**, 873–912. (Improved algorithm to compute the SVD.)

[4] Golub, G. H., and Kahan, W. (1965). "Calculating the singular values and pseudo-inverse of a matrix." *SIAM J. Numer. Anal., Ser. B*, **2**, 205–224. (Numerical method for the SVD.)

[5] Golub, G. H., and Van Loan, C. F. (1980). "An analysis of the total least squares problem." *SIAM J. Numer. Anal.*, **17**, 883–893. (Earliest analysis of TLS through SVD cited in [17].)

[6] Golub, G. H., Hoffman, A., and Stewart, G. W. (1987). "A generalization of Eckart-Young-Minski matrix approximation theorem." *Linear Algebra Appl.* **88/89**, 322-327. (One of the few publications of the E-Y-M theorem linking geometric and algebraic TLS.)

[7] Golub, G. H., and Van Loan, C. F. (1989). *Matrix Computations*, 2nd ed., John Hopkins University Press, Baltimore, MD. (Handbook of numerical linear algebra, with TLS on p. 576–581.)

[8] Grosse, E. (1995). "BibNet: a new resource for Netlib users." *SIAM News*, **28**, 4. (Occasional column on computer programs available by electronic mail.)

[9] Hogg, R. V., and Craig, A. T. (1978). *Introduction to Mathematical Statistics*, Macmillan, New York, NY.

[10] Kincaid, D., and Cheney, W. (1991). *Numerical Analysis: Mathematics of Scientific Computing*, Brooks/Cole, Pacific Grove, CA, 1991. (Mathematical background for the SVD.)

[11] Nievergelt, Y. (1994). "Exact equations for the equilibrium constants of single intermolecular complexes in terms of spectrophotometric data." *The Analyst*, **119**, 145–151. (Specific examples of Nonlinear TLS applied to find model parameters in closed form in chemistry.)

[12] Nievergelt, Y. (1994). " Total Least Squares: state-of-the-art regression in numerical analysis." *SIAM Review*, **36**, 258–264. (Short tutorial on TLS in the plane.)

[13] Nievergelt, Y. (1994). "Computing circles and spheres of arithmetic least squares." *Computer Physics Communications*, **81**, 343–350. (Fitting circles & spheres to data with Nonlinear TLS.)

[14] Stewart, G. W. (1993). "On the Early History of the Singular Value Decomposition." *SIAM Review*, **35**, 551–566. (History of the SVD.)

[15] Stoer, J., and Bulirsch, R. (1983). *Introduction to Numerical Analysis*, 2nd printing, Springer-Verlag, New York. (Mathematical background necessary to understand the algorithms.)

[16] Strang, G. (1986). *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, Wellesley, MA. (Weighting by the sample's inverse covariance matrix on p. 397–399.)

[17] Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, PA. (Comprehensive reference on TLS.)

[18] Wilkinson, J. H. (1963). *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, NJ. (Reference on digital computer's rounding errors during computations.)

[19] Wilkinson, J. H. (1988). *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, UK. (Mathematical background necessary to understand the computer algorithms.)

(GENERAL LINEAR MODEL
LATENT-ROOT REGRESSION
LEAST SQUARES
MULTIPLE LINEAR REGRESSION
NONLINEAR MODELS
NUMERICAL ANALYSIS
PRINCIPAL COMPONENT ANALYSIS
REGRESSION (VARIOUS ENTRIES)
SINGULAR-VALUE DECOMPOSITION)

YVES NIEVERGELT

satisfies all the requirements, because $S\vec{\mathbf{u}}_m = \vec{\mathbf{0}}$, which follows from $\vec{\mathbf{u}}_k^T \vec{\mathbf{u}}_m = 0$ for every $k \neq m$, and because

$$C - S = \sum_{i=1}^{m} \sigma_i \vec{\mathbf{v}} \vec{\mathbf{u}}^T - \sum_{i=1}^{m-1} \sigma_i \vec{\mathbf{v}} \vec{\mathbf{u}}^T = \sigma_m \vec{\mathbf{v}}_m \vec{\mathbf{u}}_m^T$$

has only one possibly non-zero singular value, so that $\|C - S\|_2^2 = \sigma_m^2$. □

Application 1. (Total Least Squares) Ordinary Least Squares replaces $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ and $\vec{\mathbf{b}} \in \mathbb{R}^m$ by the vector $\hat{\vec{\mathbf{b}}}$ closest to $\vec{\mathbf{b}}$ but in the column space of A. Theorems on projections then assert that $\hat{\vec{\mathbf{b}}}$ is the orthogonal projection of $\vec{\mathbf{b}}$ onto the space spanned by the columns of A, which may be computed by the factorization $A = QR$.

In contrast, Total Least Squares replaces both $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ and $\vec{\mathbf{b}} \in \mathbb{R}^m$ by a matrix $\hat{A}$ and a vector $\hat{\vec{\mathbf{b}}}$ so that $\hat{\vec{\mathbf{b}}} \in \mathcal{R}(\hat{A})$, and so that all the entries of $\hat{A}$ and $\hat{\vec{\mathbf{b}}}$ lie as closely to all the entries of A and $\vec{\mathbf{b}}$.

One way of formulating the problem just stated consists in forming the augmented matrices $C := [A; \vec{\mathbf{b}}]$ and $S := [\hat{A}; \hat{\vec{\mathbf{b}}}]$. Such matrices relate to the initial linear system by the equivalence

$$A\vec{\mathbf{x}} = \vec{\mathbf{b}} \Leftrightarrow [A; \vec{\mathbf{b}}]\,(\vec{\mathbf{x}}^T; -1)^T = \vec{\mathbf{0}}.$$

The objective of a minimum amounts to minimizing $\|C - S\|_2$, where the constraint $\hat{\vec{\mathbf{b}}} \in \mathcal{R}(\hat{A})$ amounts to imposing that the columns of S be linearly dependent, which means that S must be singular.

The theorem of Eckart, Young, & Minsky then provides a solution:

Step 1. Compute the singular value decomposition of $[A; \vec{\mathbf{b}}] = C = \sum_{i=1}^{s} \sigma_i \vec{\mathbf{v}}_i \vec{\mathbf{u}}_i^T$, with $s = \min\{p, q+1\}$.

Step 2. (Optional) Let $[\hat{A}; \hat{\vec{\mathbf{b}}}] = S = \sum_{i=1}^{s-1} \sigma_i \vec{\mathbf{v}}_i \vec{\mathbf{u}}_i^T$. Then the Total Least Squares error is $\|[A; \vec{\mathbf{b}}] - [\hat{A}; \hat{\vec{\mathbf{b}}}]\|_2 = \sigma_s = \|[A; \vec{\mathbf{b}}] - [\hat{A}; \hat{\vec{\mathbf{b}}}]\|_F$ with $[A; \vec{\mathbf{b}}] - [\hat{A}; \hat{\vec{\mathbf{b}}}] = \sigma_s \vec{\mathbf{v}}_s \vec{\mathbf{u}}_s^T$.

Step 3. The solution $(\vec{\mathbf{x}}^T; -1)^T$ is proportional to $\vec{\mathbf{v}}_s$, with -1 in the last coordinate. Therefore, $(\vec{\mathbf{x}}^T; -1)^T = (-1/[\vec{\mathbf{v}}_s]_p)\vec{\mathbf{v}}_s$.

The generic Total Least Squares problem, as just stated, has no solution if $[\vec{\mathbf{v}}_s]_p = 0$. This situation occurs if the columns of $[\hat{A}; \hat{\vec{\mathbf{b}}}]$ are linearly dependent but some column of $\hat{A}$ instead of $\hat{\vec{\mathbf{b}}}$, depends upon all the other columns.

Remark 1. $\|M\|_F$ represents the Frobenius norm of a matrix, defined by

$$\|M\|_F^2 := \sum_{k,\ell} |M_{k,\ell}|^2.$$

Because conjugation by orthogonal matrices preserves Euclidean norms, $\|UMV^T\|_F = \|M\|_F$. The singluar value decomposition of M then shows that

$$\|M\|_F^2 = \|\Sigma\|_F^2 = \sum_{k,\ell} |\Sigma_{k,\ell}|^2 = \sum_{i=1}^{n} \sigma_i^2.$$

Thus, if only one singular value differs from zero, as for $M = C - S$ above, then $\|M\|_2 = \|M\|_F$. □

Remark 2. The Total Least-Squares solution has the following geometric significance.

The rows of $[A; \vec{\mathbf{b}}]$ are data points in $\mathbb{R}^{n+1}$ but not on any hyperplane.

The rows of $[\hat{A}; \hat{\vec{\mathbf{b}}}]$ are "nearby" points on a hyperplane through $\vec{\mathbf{0}}$ and normal to $(\vec{\mathbf{x}}^T, -1)$.

The norm $\|[A; \vec{\mathbf{b}}] - [\hat{A}; \hat{\vec{\mathbf{b}}}]\|_2 = \sigma_s = \|[A; \vec{\mathbf{b}}] - [\hat{A}; \hat{\vec{\mathbf{b}}}]\|_F$ measures the sum of the squares of the distances between the given rows and the fitted rows.

If $\vec{\mathbf{b}}$ is interpreted as a function A of $\vec{\mathbf{x}}$, then the distances are measured in the projective space in homogeneous coordinates.

Definition 1. For each positive integer n, denote by $\mathcal{S}^{n-1}$ the unit sphere in $\mathbb{R}^n$. Thus,

$$\mathcal{S}^{n-1} = \left\{ \vec{\mathbf{x}} \in \mathbb{R}^n : \sum_{i=1}^{n} x_i^2 = 1 \right\}. \qquad \square$$

Lemma 1. *For all real numbers* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{n-1} > \sigma_n \geq 0$,

$$\min_{\vec{\mathbf{x}} \in \mathcal{S}^{n-1}} \sum_{i=1}^{n} (\sigma_i x_i)^2 = \min_{i \in \{1,\ldots,n\}} \sigma_i^2 = \sigma_n^2,$$

$$\max_{\vec{\mathbf{x}} \in \mathcal{S}^{n-1}} \sum_{i=1}^{n} (\sigma_i x_i)^2 = \max_{i \in \{1,\ldots,n\}} \sigma_i^2 = \sigma_1^2,$$

Proof. For each $\vec{\mathbf{x}} \in \mathcal{S}^{n-1}$, solving the equation $\sum_{i=1}^{n} x_i^2 = 1$ for x_n^2 gives $x_n^2 = 1 - \sum_{i=1}^{n-1} x_i^2$. Consequently,

$$\begin{aligned} \sum_{i=1}^{n} (\sigma_i x_i)^2 &= \sigma_n^2 x_n^2 + \sum_{i=1}^{n-1} (\sigma_i x_i)^2 \\ &= \sigma_n^2 \left(1 - \sum_{i=1}^{n-1} x_i^2 \right) + \sum_{i=1}^{n-1} \sigma_i^2 x_i^2 \\ &= \sigma_n^2 + \sum_{i=1}^{n-1} \left(\sigma_i^2 - \sigma_n^2 \right) x_i^2 \\ &\geq \sigma_n^2, \end{aligned}$$

with equality if, but only if, $x_1 = \cdots = x_{n-1} = 0$ and $x_1 = \pm 1$. The proof for the maximum is similar, with reverse inequalities, and with σ_1 singled out instead of σ_n. $\square$

Proposition 1. (Special case of the matrix approximation theorem of Eckart, Young, and Minsky.) *For all positive integers* p *and* q, *for each matrix* $C \in \mathbb{M}_{p \times q}(\mathbb{R})$, *the singular matrix* S *closest to* C *with respect to the subordinate Euclidean norm is*

$$S = \sum_{i=1}^{m-1} \sigma_i \vec{\mathbf{v}} \vec{\mathbf{u}}^T,$$

where $m := \min\{p, q\}$, *and with notation so that* C *has singular values* $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ *and singular value decomposition*

$$C = \sum_{i=1}^{m} \sigma_i \vec{\mathbf{v}} \vec{\mathbf{u}}^T = V \cdot \Sigma \cdot U^T.$$

Thus, S *results from "deleting" the smallest singular value of* C.

Proof. For each singular matrix S, a non-zero vector $\vec{\mathbf{x}}$ exists for which $\vec{\mathbf{x}} \in \mathcal{S}^{n-1}$ and $S\vec{\mathbf{x}} = \vec{\mathbf{0}}$. Expressing all vectors with respect to the orthonormal bases U in the domain $\mathbb{R}^q$ and V in the co-domain $\mathbb{R}^p$ gives, because orthogonal matrices preserve Euclidean lengths,

$$\|C - S\|_2^2 = \|V^T(C - S)U\|_2^2 = \|\Sigma - V^T S U\|_2^2 \geq \|(\Sigma - V^T S U)U^T \vec{\mathbf{z}}\|_2^2 = \|\Sigma U^T \vec{\mathbf{z}} - \vec{\mathbf{0}}\|_2^2 = \sum_{i=1}^{m} \sigma_i^2 z_i^2 \geq \sigma_m^2.$$

The inequalities just obtained mean that the minimum occurs if, but only if, $\vec{\mathbf{z}} = \pm \vec{\mathbf{u}}_m$, and then $\|C - S\|_2^2 = \sigma_m^2$. Consequently, the matrix S defined by

$$S := \sum_{i=1}^{m-1} \sigma_i \vec{\mathbf{v}} \vec{\mathbf{u}}^T$$